

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:27:20 ; Search time 8.32 Seconds
(without alignments)
150.299 Million cell updates/sec

Title: US-09-786-214A-14

Perfect score: 65

Sequence: 1 LPVVGLSPGEQE 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	67.7	381	2	F75270
2	43	66.2	821	2	C84304
3	41	63.1	156	2	F87551
4	41	63.1	326	2	T45226
5	40	61.5	243	2	I54459
6	40	61.5	428	2	AG1304
7	40	61.5	428	2	AG1676
8	40	61.5	540	2	A75250
9	40	61.5	623	2	T40991
10	39	60.0	43	2	S21065
11	39	60.0	96	2	S45441
12	39	60.0	103	2	S19975
13	39	60.0	106	2	P80070
14	39	60.0	106	2	P4282
15	39	60.0	107	2	S57444
16	39	60.0	108	2	C30502
17	39	60.0	108	2	S33988
18	39	60.0	108	2	G44151
19	39	60.0	111	2	S23628
20	39	60.0	114	2	S34095
21	39	60.0	115	1	K3HUVG
22	39	60.0	115	1	KVMSLV
23	39	60.0	115	2	S11697
24	39	60.0	116	2	B25521
25	39	60.0	119	2	S41816
26	39	60.0	125	2	PN0344
27	39	60.0	128	2	S40445
28	39	60.0	128	2	S40379
29	39	60.0	128	2	A56701

30 39 60.0 129 2 S29627
31 39 60.0 129 2 S40363
32 39 60.0 132 2 S05268
33 39 60.0 144 2 P10106
34 39 60.0 144 2 B30502
35 39 60.0 215 2 A23746
36 39 60.0 319 2 AD0941
37 39 60.0 508 2 E70764
38 39 60.0 563 2 JQ0623
39 39 60.0 1367 2 S74285
40 39 60.0 1506 2 JCS985
41 38.5 59.2 656 2 E75468
42 38 58.5 86 2 S20649
43 38 58.5 110 2 S60591
44 38 58.5 111 2 FN0537
45 38 58.5 152 2 G75184

ALIGNMENTS

RESULT 1

F75270
cytochrome P450 - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: F75270
R:White, O.; Eissen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75270
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-381 <WHI>
A:Cross-references: GB:AE02076; GB:AE000513; NID:96460285; PIDN:AAF12016.1; PID:964602
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2473
A:Map position: 1

Query Match 67.7% Score 44; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 8.7; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0

Qy 1 LPVVVGLSP 9
|||
Db 52 LPVVVGLSP 60

RESULT 2

C84304
DNA helicase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84304
R:NG, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky,
Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L.
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: C84304
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-821 <STO>
A:Cross-references: GB:AE004437; NID:g10580995; PIDN:AAG19799.1; GSPDB:GN00138
C:Genetics:
A:Gene: hel

```

Query Match      66.2%; Score 43; DB 2; Length 821;
Best Local Similarity 90.0%; Pred. No. 28;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AVVGLSPGQE 12
Db 326 AVVGLSPAEQ 335

RESULT 3
F87551
conserved hypothetical protein CC2439 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: F87551
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87551
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <STO>
A:Cross-references: GB:AE005673; NID:gl3423984; PIDN:AAK24410.1; GSPDB:GN00148
C:Genetics:
C:Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

Query Match      63.1%; Score 41; DB 2; Length 156;
Best Local Similarity 80.0%; Pred. No. 12;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AVVGLSPGQE 12
Db 18 AVVGLDPGEK 27

RESULT 4
T45226
probable N5,N10-methylene-tetrahydrodethanopterin reductase (F420-dependent) [imported]
C:Species: Methanobolus tindarius
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45226
R:Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G.;
submitted to the EMBL Data Library, September 1998
A:Description: The F420H2-dehydrogenase from Methanobolus tindarius: Cloning of the ffd
A:Reference number: Z22947
A:Accession: T45226
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-326 <WES>
A:Cross-references: EMBL:AJ011519; PIDN:CAB56639.1
A:Experimental source: DSM 2278
C:Genetics:
A:Gene: ffdA

Query Match      63.1%; Score 41; DB 2; Length 326;
Best Local Similarity 70.0%; Pred. No. 24;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AVVGLSPGQE 12
Db 88 AILGLGPGQE 97

RESULT 5
I54459
MHC H-2K1-k - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C:Accession: I54459

```

```

R:Watts, S.; Davis, A.C.; Goodenow, R.S.
Immunogenetics 29, 355-357, 1989
A:Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.
A:Reference number: I54459; MUID:89233303; PMID:2714856
A:Accession: I54459
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-243 <RES>
A:Cross-references: GB:M27134; NID:gl199435; PIDN:AAA39610.1; PID:g387456
C:Genetics:
C:Introns: 22/1; 112/1

Query Match      61.5%; Score 40; DB 2; Length 243;
Best Local Similarity 77.8%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VGLSPGQE 13
Db 216 LGLSPGEE 224

RESULT 6
AG1304
uracil permease homolog pyrP [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AG1304
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wetland
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AG1304
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC99917.1; PID:gl16411293; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: pyrP
C:Superfamily: uracil transport protein uraA

Query Match      61.5%; Score 40; DB 2; Length 428;
Best Local Similarity 70.0%; Pred. No. 47;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPVVGGLSPG 10
Db 42 VPSVTGLSPG 51

RESULT 7
AG1676
uracil permease homolog pyrP [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AG1676
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wetland
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AG1676
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97183.1; PID:gl16414454; GSPDB:GN00178

```

A;Experimental source: strain Clip11262

C;Genetics:

A;Gene: pyrP

C;Superfamily: uracil transport protein uraA

Query Match 61.5%; Score 40; DB 2; Length 428;

Best Local Similarity 70.0%; Pred. No. 47;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAVVGSLSPG 10

Db 42 VPSVTGLSPG 51

RESULT 8

A75250

carboxylesterase, type B - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C;Accession: A75250

R;White, O.; Eissen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: A75250

A;Molecule type: DNA

A;Residues: 1-540 <WHI>

A;Cross-references: GB:AE002092; GB:AE000513; NID:g6460455; PIDN:AAF12163.1; PID:g646045

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR2626

A;Map position: 1

C;Superfamily: cholinesterase; cholinesterase homology

Query Match 61.5%; Score 40; DB 2; Length 540;

Best Local Similarity 70.0%; Pred. No. 60;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PAVVGLSPGE 11

Db 512 PQVGLTAPGE 521

RESULT 9

T40991

probable lysophospholipase precursor - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000

C;Accession: T40991

R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.

submitted to the EMBL Data Library, March 1999

A;Reference number: Z21962

A;Accession: T40991

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-623 <LYN>

A;Cross-references: EMBL:AL049559; PIDN:CAR40176.1; GSPDB:GN000068; SPDB:SPCC1450.09C

A;Experimental source: strain 972h; cosmid c1450

C;Genetics:

A;Gene: SPDB:SPCC1450.09C

A;Map position: 3

C;Superfamily: yeast lysophospholipase

Query Match 61.5%; Score 40; DB 2; Length 623;

Best Local Similarity 75.0%; Pred. No. 69;

Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 PAVVGLSPGE 13

Db 76 PASDGLSTGE 87

RESULT 10

S21065

Ig kappa chain V region (anti-RH(D)) - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 22-Nov-1993 #sequence_revision 24-May-1996 #text_change 09-May-1997

C;Accession: S21065

R;Plouha, A.; Lecroisey, A.; Henschen, A.; Rouger, P.; Keil, B.

Protein Seq. Data Anal. 4, 317-318, 1991

A;Title: Subgroup assignment of a human monoclonal anti-Rh(D) antibody.

A;Reference number: S21065; MUID:92253544; PMID:1812483

A;Accession: S21065

A;Molecule type: protein

A;Residues: 1-43 <DLO>

C;Superfamily: immunoglobulin V region; immunoglobulin homology

C;Keywords: heterotetramer; immunoglobulin

Query Match 60.0%; Score 39; DB 2; Length 43;

Best Local Similarity 63.6%; Pred. No. 69;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGE 12

Db 8 PATLSLSPGER 18

RESULT 11

S45441

Ig kappa chain V region - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jan-2000

C;Accession: S45441

R;Cox, J.P.L.; Tomlinson, I.M.; Winter, G.

Eur. J. Immunol. 24, 827-836, 1994

A;Title: A directory of human germ-line V(kappa) segments reveals a strong bias in their

A;Reference number: S45324; MUID:94200218; PMID:8149953

A;Accession: S45441

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-96 <COX>

A;Cross-references: EMBL:Z27500

C;Superfamily: immunoglobulin V region; immunoglobulin homology

C;Keywords: heterotetramer; immunoglobulin

F;16-91/Domain: immunoglobulin homology <IMM>

Query Match 60.0%; Score 39; DB 2; Length 96;

Best Local Similarity 63.6%; Pred. No. 15;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGE 12

Db 8 PATLSLSPGER 18

RESULT 12

S19375

Ig kappa chain V region (M-T408) - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jan-2000

C;Accession: S19375

R;Weissenhorn, W.; Riethmuller, G.; Weiss, E.M.; Rieber, E.P.

submitted to the EMBL Data Library, March 1992

A;Description: Structural characterization of CD4 mAb.

A;Reference number: S19363

A;Accession: S19375

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-103 <WFI>

A;Cross-references: EMBL:X65097; NID:g52296; PIDN:CAA46225.1; PID:g52297

C;Superfamily: immunoglobulin V region; immunoglobulin homology

C;Keywords: heterotetramer; immunoglobulin

F;11-85/Domain: immunoglobulin homology <IMM>

Query Match 60.0%; Score 39; DB 2; Length 103;
 Best Local Similarity 63.6%; Pred. No. 17;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 PAVVGLSPGEQ 12
 ||:|||||:
 Db 3 PATLSLSPGER 13

RESULT 13

Ig kappa chain V region (38C13.V6.1) - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 21-Jan-2000
 C:Accession: PS0070
 R;Levy, S.; Campbell, M.J.; Levy, R.
 J. Exp. Med. 170, 1-13, 1989
 A;Title: Functional immunoglobulin light chain genes are replaced by ongoing rearrangement
 A;Reference number: A92781; MUID:89310348; PMID:2501443
 A;Accession: PS0070
 A;Status: translation not shown
 A;Molecule type: DNA
 A;Residues: 1-106 <LEV>
 C;Superfamily: immunoglobulin V region; immunoglobulin homology
 C;Keywords: heterotetramer; immunoglobulin
 F;16-89/Domain: immunoglobulin homology <IMM>

Query Match 60.0%; Score 39; DB 2; Length 106;
 Best Local Similarity 54.5%; Pred. No. 17;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
 ||:|||||:
 Db 8 PATIAASPGEX 18

RESULT 14

PC4282
 Ig kappa chain (anti-SS-A/Ro 60K peptide) (E-42 and E-56) - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 21-Jan-2000
 C:Accession: PC4282; PC4284
 R;Suzuki, H.; Takemura, H.; Suzuki, M.; Sekine, Y.; Kashiwagi, H.
 Biochem. Biophys. Res. Commun. 232, 101-106, 1997
 A;Title: Molecular cloning of anti-ss-A/Ro 60-kDa peptide fab fragments from infiltrating
 A;Reference number: PC4279; MUID:97236289; PMID:9125110
 A;Accession: PC4282
 A;Molecule type: protein
 A;Residues: 1-106 <SU2>
 A;Note: E-42
 A;Accession: PC4284
 A;Molecule type: protein
 A;Residues: 1-106 <SU2>
 A;Note: E-56
 C;Comment: This antibody is commonly found in systemic autoimmune diseases such as Sjogren
 C;Superfamily: immunoglobulin V region; immunoglobulin homology
 F;14-88/Domain: immunoglobulin homology <IMM>

Query Match 60.0%; Score 39; DB 2; Length 106;
 Best Local Similarity 63.6%; Pred. No. 17;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
 ||:|||||:
 Db 6 PATLSLSPGER 16

RESULT 15

SS7444
 Ig kappa chain V-J region - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 10-Oct-1995 #sequence_revision 17-Nov-1995 #text_change 21-Jan-2000
 C:Accession: SS7444

R;Paterson, G.; Wilson, G.; Kennedy, P.G.E.; Willison, H.J.

submitted to the EMBL Data Library, June 1995

A;Description: Analysis of anti-GM1 ganglioside IgM antibodies cloned from motor neurop

A;Reference number: SS7408

A;Accession: SS7444

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-107 <PAT>

A;Cross-references: EMBL:X87898; NID:g871275; PIDN:CAA61149.1; PID:g871276

C;Superfamily: immunoglobulin V region; immunoglobulin homology

F;16-90/Domain: immunoglobulin homology <IMM>

Query Match 50.0%; Score 39; DB 2; Length 107;
 Best Local Similarity 63.6%; Pred. No. 17;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
 ||:|||||:
 Db 8 PATLSLSPGER 18

Search completed: May 7, 2004, 12:39:07
 Job time : 8.48667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 4.68 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-13

Perfect score: 68

Sequence: 1 PAVVGLSPGEQY 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	63.2	211	1 COBL_METJA	Q58917 methanococ
2	43	63.2	633	1 PLB5_SCHPO	Q9y7h6 schizosacch
3	41	60.3	156	1 RUVX_CAUCR	Q9a5k8 caulobacter
4	41	60.3	326	1 MER_METTI	Q9uxp0 methanolobu
5	41	60.3	508	1 COBI_MYCTU	Q10677 mycobacteri
6	40	58.8	673	1 PLB4_SCHPO	Q9p327 schizosacch
7	39	57.4	115	1 KV31_HUMAN	P04433 homo sapien
8	39	57.4	115	1 KV51_MOUSE	P01642 mus musculu
9	39	57.4	1636	1 BUD3_YEAST	P25558 saccharomyc
10	38	55.9	429	1 R51_LEULA	P50889 leuconostoc
11	38	55.9	507	1 CAT3_PICAN	P30263 pichia angu
12	37	54.4	446	1 COB3_ARCFU	O29534 a cobalamin
13	37	54.4	446	1 ENO1_MALZE	P26301 zea mays (m
14	37	54.4	597	1 NR41_RAT	P22829 rattus norv
15	37	54.4	607	1 GLM5_CLOTE	Q990u2 c glucosami
16	37	54.4	813	1 CADM_MOUSE	Q9wtp5 mus musculu
17	37	54.4	1402	1 N160_MOUSE	Q9z0w3 mus musculu
18	36.5	53.7	1121	1 BMS1_SCHPO	O94653 schizosacch
19	36	52.9	129	1 KV3H_HUMAN	P04207 homo sapien
20	36	52.9	158	1 RSD_ECOLI	P31690 escherichia
21	36	52.9	162	1 RSD_SALTY	Q91916 salmonella
22	36	52.9	243	1 SMT2_SYNP7	P42451 synechococc
23	36	52.9	279	1 TYSY_CAUCR	Q9a6h0 caulobacter
24	36	52.9	317	1 OAD1_HUMAN	Q9nge0 homo sapien
25	36	52.9	342	1 HUPK_AZOVI	P31878 azotobacter
26	36	52.9	348	1 HOKV_AZOVI	P40597 azotobacter
27	36	52.9	422	1 TKSU_PYRKO	P58502 pyrococcus
28	36	52.9	464	1 LEU2_BACCR	Q81g10 bacillus ce
29	36	52.9	471	1 CD36_BOVIN	P26201 bos taurus
30	36	52.9	590	1 MUTL_THETN	Q8a70 thermoanaer
31	36	52.9	609	1 GLM5_CLOPE	Q8xh27 c glucosami
32	36	52.9	614	1 CPRI_DROME	Q9vn93 drosophila
33	36	52.9	624	1 PLB2_SCHPO	O13857 schizosacch

34	36	52.9	699	1 EFG_HABIN	P43925 haemophilus
35	36	52.9	700	1 EFG_PASMU	P57938 pasteurella
36	36	52.9	750	1 ELS_CHICK	P07916 gallus gall
37	36	52.9	844	1 HEXA_STRPN	P10564 streptococ
38	36	52.9	926	1 PTM4_HUMAN	P29074 homo sapien
39	36	52.9	1103	1 VG37_BPARI	Q990b5 bacterioph
40	36	52.9	1173	1 ATC2_YEAST	P38929 saccharomyc
41	36	52.9	3703	1 ABF1_HUMAN	Q15911 homo sapien
42	35	51.5	100	1 KV3C_HUMAN	P01621 homo sapien
43	35	51.5	109	1 KV3B_HUMAN	P01620 homo sapien
44	35	51.5	109	1 KV3D_HUMAN	P01622 homo sapien
45	35	51.5	109	1 KV3E_HUMAN	P01623 homo sapien

ALIGNMENTS

RESULT 1
COBL_METJA STANDARD; PRT; 211 AA.
ID COBL_METJA STANDARD; PRT; 211 AA.
AC Q58917;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable precorrin-6Y C5,15-methyltransferase [decarboxylating]
DE (BC 2.1.1.132) (Precorrin-6 methyltransferase) (Precorrin-6Y
DE methylase).
DE COBL OR MJ1522.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8689087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
Scott J.L., Geoghagen N.S.M., Weidman J.F., Sadow P.W., Hanna M.C.,
Utterback T.R., Kelley J.M., Peterson J.D., Haurst M.A., Kaine B.P., Borodovsky M.,
Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
*Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii*;
Science 273:1058-1073 (1996).
RL -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
PRECORRIN-6Y TO FORM PRECORRIN-8X (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: TO S-TYPEHIMURIUM CBIE; ALSO, LOW, TO OTHER
METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U67593; AAB99541.1; -
DR PIR; A64490; A64490.
DR TIGR; MJ1522; -
DR InterPro; IPR000878; Cor/por Mettransf.
DR Pfam; PF00590; Tr_methylase; 1.
DR Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 211 AA; 23805 MW; 279A1A2B14369510 CRC64;

Query Match 63.2%; Score 43; DB 1; Length 211;

Best Local Similarity 54.5%; Pred. No. 3.8;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVGLSPGQEQY 13
Db 4 IVGIGPGDREY 14

RESULT 2

PLB5 SCHPO STANDARD; PRT; 633 AA.
ID Q917N6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Putative lysophospholipase C1450.09c precursor (EC 3.1.1.5)
DE (Phospholipase B).
GN SPC1450.09c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_taxID=4996;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21849401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.M.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Fumelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomycetes pombe.";
Nature 415:871-880 (2002).
CC -!- FUNCTION: Catalyzes the release of fatty acids from
CC lysophospholipids (By similarity).
CC -!- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC glycerophosphocholine + a fatty acid anion.
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- SIMILARITY: Belongs to the lysophospholipase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL049559; CAB40176.2; -;
CC GenBank; SPombe; SPC1450.09c; -;
CC InterPro; IPR002642; PLAC.
CC Pfam; PF01735; PLA2_B; 1.
CC SMART; SM00022; PLAC; 1.

Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 633 PUTATIVE LYSOPHOSPHOLIPASE C1450.09c.
FT CARBOHYD 118 118 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 508 508 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 537 537 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 633 AA; 68292 MW; 49871B2955893D19 CRC64;

Query Match 63.2%; Score 43; DB 1; Length 633;
Best Local Similarity 69.2%; Pred. No. 11;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAVVGLSPGQEQY 13
Db 76 PASDGLSTGQEQF 88

RESULT 3

RVUX CAUCR STANDARD; PRT; 156 AA.
ID RVUX CAUCR
AC Q9A5K8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Putative Holliday junction resolvase (EC 3.1.1.-).
GN CC2439.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_taxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CBJ5;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feidilyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Ely B.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uitterlbeck T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro B., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).
RL -!- FUNCTION: Could be a nuclease that resolves Holliday junction
CC intermediates in genetic recombination.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the YGF HJR family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC


```

CC -----
DR EMBL; Z73966; CAA98214.1; -.
DR EMBL; AE007063; AA46406.1; -.
DR EMBL; BX248341; CAD96945.1; -.
DR TIGR; E70764; E70764.
DR TIGR; MT2126; -.
DR TubercuList; RV2066; -.
DR InterPro; IPR006364; Cobi_CbIL.
DR InterPro; IPR006363; Cobi_CbIL.
DR InterPro; IPR000878; Cor/por_Mettransf.
DR InterPro; IPR003043; Uropor_Mettransf.
DR Pfam; PF00590; TP_methylase; 2.
DR TIGRfams; TIGR01467; cobi_cbiL; 1.
DR TIGRfams; TIGR01466; cobi_cbiH; 1.
DR PROSITE; PS00839; SUMT_1; 1.
DR PROSITE; PS00840; SUMT_2; 1.
DR COBALANIN biosynthesis; Porphyryn biosynthesis; Transferase;
KW Cobalamin biosynthesis; Complete proteome.
KW Methyltransferase; Multifunctional enzyme; Complete proteome.
FT DOMAIN 1 243 PRECORRIN-2 C20-METHYLTRANSFERASE.
FT DOMAIN 244 508 PRECORRIN-3 METHYLASE.
FT SEQUENCE 508 AA; 53910 MW; 95AC066F022C4DC1 CRC64;
SQ
Query Match 60.3%; Score 41; DB 1; Length 508;
Best Local Similarity 58.3%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 AVVGLSPGEQY 13
||||| |||: ::
DB 250 AVVGLGDSWD 261

RESULT 6
PLB4_SCHPO
ID PLB4_SCHPO STANDARD; PRT; 673 AA.
AC Q9P327;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative lysophospholipase C977.09c precursor (EC 3.1.1.5)
DE (Phospholipase B).
GN SPAC977.09C OR SPAC1348.10C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=2184401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares S., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grynoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Fohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Furnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

```

```

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415: 871-880(2002).
CC -|- FUNCTION: Catalyzes the release of fatty acids from
CC lysophospholipids (By similarity).
CC -|- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC glycerophosphocholine + a fatty acid anion.
CC -|- SUBCELLULAR LOCATION: Secreted (Probable).
CC -|- SIMILARITY: Belongs to the lysophospholipase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL358912; CAB94277.1; -.
CC EMBL; AL37130; CAB9631.1; -.
CC PIR; T50281; T50281.
CC GeneDB SPombe; SPAC977.09c; -.
CC InterPro; IPR001179; FKBP_PPIase.
CC InterPro; IPR002642; PLAC.
CC Pfam; PF01735; PLA2_B; 1.
CC SMART; SMO0022; PLAC; 1.
CC KW Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 673 PUTATIVE LYSOPHOSPHOLIPASE C977.09C.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 516 516 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 574 574 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 673 AA; 74595 MW; B39A77E76CD694B CRC64;
Query Match 58.8%; Score 40; DB 1; Length 673;
Best Local Similarity 61.5%; Pred. No. 39;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQY 13
||||| |||: |||
DB 83 PASEGLNEGEQSY 95

RESULT 7
KV31_HUMAN
ID KV31_HUMAN STANDARD; PRT; 115 AA.
AC P04433;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-III region VG precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85087932; PubMed=6440122;
RA Pech M., Zachau H.G.;
RT "Immunoglobulin genes of different subgroups are interdigitated
RT within the VK locus";
RL Nucleic Acids Res. 12:9229-9236(1984).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X01668; -; NOT_ANNOTATED_CDS.
 DR PIR; A01900; K3HUVG.
 DR HSP; P80362; 1MTL.
 DR GO; GO:0005576; C:extracellular; NAS.
 DR GO; GO:0003823; F:antigen binding; NAS.
 DR GO; GO:0006955; P:immune response; NAS.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig; 1.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS05835; IG LIKE; 1.
 KW Immunoglobulin V region; Signal.

FT SIGNAL 1 20 IG KAPPA CHAIN V-III REGION VG.
 FT CHAIN 21 >115 FRAMEWORK-1.
 FT DOMAIN 21 43 COMPLEMENTARITY-DETERMINING-1.
 FT DOMAIN 44 54 FRAMEWORK-2.
 FT DOMAIN 55 69 COMPLEMENTARITY-DETERMINING-2.
 FT DOMAIN 70 76 FRAMEWORK-3.
 FT DOMAIN 77 108 COMPLEMENTARITY-DETERMINING-3.
 FT DOMAIN 109 115 BY SIMILARITY.
 FT DISULFID 43 108
 FT NON TER 115 115
 SQ SEQUENCE 115 AA; 12575 MW; 2DE47CDA3A17D555 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 115;
 Best Local Similarity 63.6%; Pred. No. 10;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQ 11
 DB 28 PAVLSVSPGER 38

RESULT 8
 KVS1_MOUSE
 ID KV51_MOUSE STANDARD; PRT; 115 AA.
 AC P01642;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig kappa chain V-V region L7 precursor (Fragment).
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81220975; PubMed=6264318;
 RA Pech M., Hochtl J., Schnell H., Zachau H.G.;
 RT "Differences between germ-line and rearranged immunoglobulin V kappa
 RT coding sequences suggest a localized mutation mechanism."
 RL Nature 291:668-670(1981).
 CC -1- MISCELLANEOUS: THERE APPEAR TO BE TWO POSSIBLE SPLICE JUNCTIONS AT
 CC THE 3' END OF THE INTRON. THE ALTERNATE WOULD CODE FOR A PROTEIN
 CC LACKING RESIDUES 17-19.

DR PIR; A01925; KVM5L7.
 DR PDB; 1J10; 18-FEB-03.
 DR PDB; 1J1P; 18-FEB-03.
 DR PDB; 1J1X; 18-FEB-03.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003596; Ig_V.
 DR Pfam; PF00047; Ig; 1.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS05835; IG LIKE; 1.

KW Immunoglobulin V region; Signal; 3D-structure.

FT SIGNAL 1 20
 FT CHAIN 21 >115 IG KAPPA CHAIN V-V REGION L7.
 FT DOMAIN 21 43 FRAMEWORK-1.
 FT DOMAIN 44 54 COMPLEMENTARITY-DETERMINING-1.
 FT DOMAIN 55 69 FRAMEWORK-2.
 FT DOMAIN 70 76 COMPLEMENTARITY-DETERMINING-2.
 FT DOMAIN 77 108 FRAMEWORK-3.
 FT DOMAIN 109 >115 COMPLEMENTARITY-DETERMINING-3.
 FT DISULFID 43 108 BY SIMILARITY.
 FT NON TER 115 115
 SQ SEQUENCE 115 AA; 12615 MW; C17BEC758C577E00 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 115;
 Best Local Similarity 54.5%; Pred. No. 10;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQ 11
 DB 28 PAVLSVSPGER 38

RESULT 9

BUD3_YEAST
 ID BUD3_YEAST STANDARD; PRT; 1636 AA.
 AC P25558; P25556; P25557; P87007;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bud site selection protein BUD3.
 DE BUD3 OR YCL014W/YCL013W/YCL012W OR YCL14W/YCL13W/YCL12W.
 GN Saccharomyces cerevisiae (Baker's yeast).
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95247824; PubMed=7730410;
 RA Chant J., Mischke M., Mitchell E., Herskowitz I., Pringle J.R.;
 RT "Role of Bud3p in producing the axial budding pattern of yeast."
 RL J. Cell Biol. 129:767-778(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=92244356; PubMed=1574125;
 RA Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
 RA Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
 RA Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
 RA Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
 RA Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
 RA De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
 RA Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
 RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
 RA Frangoules-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
 RA Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Glandsdorff N.,
 RA Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haasemann M.,
 RA Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
 RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
 RA Jacq C., Jaquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
 RA Kleinhans U., Kreis P., Lafranchi G., Lewis C., van der Linden C.G.,
 RA Lucchini G., Lutzenkirchen K., Maat C., Mannheim G., Manzano M.E.,
 RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., Navas L.,
 RA Messenguy F., Mewes H.-W., Molenans F., Montague M.A., Navas L.,
 RA Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
 RA Perea J., Philippsen P., Pierard A., Planta R.J., Plevani S.,
 RA Poetsch B., Pohl F.M., Purnelle B., Richterich P., Roberts A.B., Rodriguez F.,
 RA Raynal A., Remacha M., Richterich P., Sor F., Soustelle C.,
 RA Shu Y., Skala J., Slonimski P.P., Sor F., Steiner S., Thierry A.,
 RA Spiegelberg R., Staveva L.I., Steensma H.Y., Steiner S., Vetter I.,
 RA Thireos G., Triano L.N., Urrestazu L.A., Valle G., Vetter I.,
 RA van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warming J.R.,
 RA von Wettstein D., Wickstead B.L., Wilson C., Wurst H., Xu G.,

```

RA Zimmermann F.K., Sgouros J.G.;
RT "The complete DNA sequence of yeast chromosome III.";
RL Nature 357:38-46(1992).
[3]
RP REVISIONS.
RA Gromadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
[4]
RN REVISIONS.
RA Valles G., Volckaerts G.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Co-assembles with BUD4 at bud sites. BUD4 and BUD3 may
CC cooperate to recognize a spatial landmark (the neck filaments)
CC during mitosis and they subsequently become a landmark for
CC establishing the axial budding pattern in G1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U17580; AAA86315.1; -.
CC EMBL; X59720; CAM42346.2; -.
CC PIR; S74285; S74285.
CC PIR; S74286; S74286.
CC GeneOnline; 138855; -.
CC SGD; S0000520; BUD3.
CC InterPro; IPR000219; RhoGEF.
CC SMART; SM00325; RhoGEF; 1.
CC Cell cycle.
KW
SQ SEQUENCE 1636 AA; 184717 MW; 9E4E46BA5C3A3F69 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 1636;
Best Local Similarity 58.3%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAVVGLSPGGEQ 12
||: |||: |||
Db 112 PAIENLSPDQEF 123

RESULT 10
RS1 LEULA STANDARD; PRT; 429 AA.
AC P50889; P71450;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 40S ribosomal protein S1.
GN RPS1.
OS Leuconostoc lactis.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1246;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97186703; PubMed=9034319;
RA Yamit-Hezi A., Levy Z., Neuman S., Nudel U.;
RT "A Leuconostoc lactis protein with homology to ribosomal protein S1
RT shares common epitopes and common DNA binding properties with a
RT mammalian DNA binding nuclear factor.";
RL Gene 185:99-103(1997).
[2]
RN [2]
RP SEQUENCE OF 24-429 FROM N.A.
RX MEDLINE=95237615; PubMed=7721096;
RA Eklund E.A., Lee S.W., Skalniak D.G.;
RT "Cloning of a cDNA encoding a human DNA-binding protein similar to
RT ribosomal protein S1.";
RL Gene 155:231-235(1995).
[3]
RN [3]
RP SEQUENCE OF 78-429 FROM N.A.

```

```

RX MEDLINE=96164600; PubMed=8568274;
RA Tsuzaka K., Ieu A.K., Frank M.B., Movafagh B.F., Koscec M.,
RA Winkler T.H., Kalden J.R., Reichlin M.;
RT "Lupus autoantibodies to double-stranded DNA cross-react with
RT ribosomal protein S1.";
RL J. Immunol. 156:1668-1675(1996).
CC -!- FUNCTION: EXHIBITS PREFERENTIAL BINDING TO SINGLE-STRANDED AND
CC DOUBLE-STRANDED DNA AND A LOW BINDING AFFINITY FOR RNA.
CC -!- SIMILARITY: Belongs to the S1P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 4 S1 motif domains.
CC -!- CAUTION: WAS ORIGINALLY (REF.2 AND REF.3) THOUGHT TO ORIGINATE
CC FROM HUMAN BUT IS MOST PROBABLY THE RESULT OF A CDNA LIBRARY
CC CONTAMINATION BY L.LACTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U24086; AAB08978.1; -.
CC EMBL; U05589; AAA77669.1; -.
CC EMBL; U27517; AAA97575.1; -.
CC HSP; P05055; ISRO.
CC InterPro; IPR008994; Nucleic_acid OB.
CC InterPro; IPR000110; Ribosomal_S1.
CC InterPro; IPR003029; S1.
CC Pfam; PF00575; S1; 4.
CC PRINTS; PR00681; RIBOSOMALS1.
CC SMART; SM00316; S1; 4.
CC PROSITE; PS50126; S1; 4.
CC KW Ribosomal protein; Repeat; RNA-binding.
FT DOMAIN 55 128 S1 MOTIF 1.
FT DOMAIN 144 211 S1 MOTIF 2.
FT DOMAIN 231 299 S1 MOTIF 3.
FT DOMAIN 316 385 S1 MOTIF 4.
FT CONFLICT 24 24 S -> G (IN REF. 2).
FT CONFLICT 122 122 A -> S (IN REF. 3).
FT CONFLICT 217 217 L -> R (IN REF. 2 AND 3).
SQ SEQUENCE 429 AA; 46386 MW; 92AC82605F39DDEC CRC64;

Query Match 55.9%; Score 38; DB 1; Length 429;
Best Local Similarity 80.0%; Pred. No. 55;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 AVVGLSPGEQ 11
|||||
Db 71 AVVGLSTGEE 80

RESULT 11
CATA PICAN STANDARD; PRT; 507 AA.
ID CATA PICAN STANDARD; PRT; 507 AA.
AC P30263;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peroxisomal catalase (BC 1.11.1.6).
GN PXP9 OR PXP-9.
OS Pichia angusta (Yeast) (Hansenula polymorpha).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4905;
RN [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 34438;
RX MEDLINE=92299073; PubMed=1607006;
RA Didion T., Roggenkamp R.O.;
RT "Targeting signal of the peroxisomal catalase in the methylotrophic
RT yeast Hansenula polymorpha.";
RL FEBS Lett. 303:113-116(1992).

```

CC -!- FUNCTION: Occurs in almost all aerobically respiring organisms and
 CC serves to protect cells from the toxic effects of hydrogen
 CC peroxide.
 CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
 CC -!- COFACTOR: Heme group.
 CC -!- SUBUNIT: Homotrimer.
 CC -!- SUBCELLULAR LOCATION: Peroxisomal.
 CC -!- SIMILARITY: Belongs to the catalase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X56501; CAA39856.1; -.
 CC PIR; S23422; S23422.
 CC HSSP; P15202; IA4E.
 CC InterPro; IPR002226; Catalase.
 CC Pfam; PF001199; Catalase; 1.
 CC PRINTS; PR00067; CATALASE.
 CC ProDom; PD000510; Catalase; 1.
 CC PROSITE; PS00437; CATALASE 1; 1.
 CC PROSITE; PS00438; CATALASE 2; 1.
 CC Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
 CC Peroxisome.
 CC ACT_SITE 65 65 BY SIMILARITY.
 CC FT ACT_SITE 138 138 BY SIMILARITY.
 CC FT METAL 348 348 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 CC FT SITE 505 507 MICROBODY TARGETING SIGNAL (POTENTIAL).
 CC SEQUENCE 507 AA; 57849 MW; 3536ED0A49539CC3 CRC64;
 CC -----
 CC Query Match 55.9%; Score 38; DB 1; Length 507;
 CC Best Local Similarity 70.0%; Pred. No. 65;
 CC Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 3 VVGLSPGGEQ 12
 CC | : : : : :
 CC Db 445 VLGRTPGGEQ 454
 CC -----
 CC RESULT 12
 CC COBJ ABCFU
 CC ID COBJ_ARCFU STANDARD; PRT; 446 AA.
 CC AC O29534;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Cobalamin biosynthesis protein cobJH [includes: Precorrin-3B C17-
 CC methyltransferase [EC 2.1.1.131] (Precorrin-3 methyltransferase)
 CC DE (Precorrin-3 methylase); Precorrin-8X methylmutase (EC 5.4.1.2)
 CC DE (Precorrin isomerase)].
 CC COBJH OR AF0724.
 CC GN COBJH
 CC OS Archaeoglobus fulgidus.
 CC OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 CC OC Archaeoglobaceae; Archaeoglobus.
 CC OX NCBI_TaxID=2234;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 CC MEDLINE=98049343; PubMed=989475;
 CC RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 CC Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 CC Richardson D.L., Kierlavage A.R., Graham D.B., Kyripides N.C.,
 CC Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 CC Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 CC Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 CC Overbeek R., Gayney J.D., Weidman J.F., McDonald L., Uutterback T.,
 CC Cotton M.D., Spriggs T., Artlach P., Kaine B.P., Sykes S.M.,
 CC Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 CC Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,

RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RL reducing archaeon *Archaeoglobus fulgidus*.";
 CC -!- FUNCTION: BIFUNCTIONAL ENZYME WITH A METHYLTRANSFERASE DOMAIN THAT
 CC CATALYZES THE METHYLATION OF C-17 IN PRECORRIN-3B TO FORM
 CC PRECORRIN-4 AND AN ISOMERASE DOMAIN THAT CATALYZES THE CONVERSION
 CC OF PRECORRIN-8X TO HYDROGENOBYRINIC ACID; A METHYL MIGRATION
 CC REACTION DURING THE TRANSFORMATION OF PRECORRIN-3 TO FORM
 CC COBYRINIC ACID (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-3B = S-
 CC adenosyl-L-homocysteine + precorrin 4.
 CC -!- CATALYTIC ACTIVITY: Precorrin-8X = hydrogenobyrinate.
 CC -!- PATHWAY: Cobalamin biosynthesis.
 CC -!- SIMILARITY: Belongs to the cobH family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001055; AAE90518.1; -.
 CC PIR; D69340; D69340.
 CC TIGR; AF0724; -.
 CC InterPro; IPR003722; CbiC.
 CC InterPro; IPR006363; CobJ.
 CC InterPro; IPR000878; Cor/por_Mettransf.
 CC Pfam; PF02570; CbiC; 1.
 CC Pfam; PF00590; TP_methylase; 1.
 CC TIGRFAMS; TIGR01456; cobJ cbiH; 1.
 CC Cobalamin biosynthesis; Transferase; Methyltransferase; Isomerase;
 CC Multifunctional enzyme; Complete proteome.
 CC FT DOMAIN 1 246 PRECORRIN-3B C-17 METHYLTRANSFERASE.
 CC FT DOMAIN 247 446 PRECORRIN-8X METHYLMUTASE.
 CC SEQUENCE 446 AA; 48678 MW; 7341BCBC0998FFDA CRC64;
 CC -----
 CC Query Match 54.4%; Score 37; DB 1; Length 446;
 CC Best Local Similarity 60.0%; Pred. No. 85;
 CC Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 3 VVGLSPGGEQ 12
 CC | : : : : :
 CC Db 12 VVG:GPGKEE 21
 CC -----
 CC RESULT 13
 CC ENOI MAIZE
 CC ID ENOI_MAIZE STANDARD; PRT; 446 AA.
 CC AC P26301;
 CC DT 01-MAY-1992 (Rel. 22, Created)
 CC DT 01-MAY-1992 (Rel. 22, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Enolase 1 (EC 4.2.1.11) (2-phosphoglycerate dehydratase 1) (2-phospho-
 CC D-glycerate hydro-lyase 1).
 CC GN ENOI OR PGH1.
 CC OS Zea mays (Maize).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 CC OX NCBI_TaxID=4577;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=cv. Berkeley Fast; TISSUE=Root;
 CC MEDLINE=91316216; PubMed=1859865;
 CC RA Lal S.K., Johnson S., Conway T., Kelley P.M.,
 CC "Characterization of a maize cDNA that complements an enolase-
 CC deficient mutant of *Escherichia coli*.";
 CC Plant Mol. Biol. 16:787-795 (1991).
 CC -!- CATALYTIC ACTIVITY: 2-phospho-D-glycerate = phosphoenolpyruvate +
 CC H(2)O.

CC -|- COFACTOR: Magnesium is required for catalysis and for stabilizing
 CC the dimer (by similarity).
 CC -|- PATHWAY: Glycolysis.
 CC -|- SUBUNIT: Homodimer.
 CC -|- SIMILARITY: Belongs to the enolase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55981; CAA39454.1; --
 DR PIR; S16257; S16257.
 DR HSSP; P56252; 1PDZ.
 DR MaizelDB; 30060; --
 DR InterPro; IPR000941; Enolase.
 DR Pfam; PF00113; enolase_1.
 DR Pfam; PF03952; enolase_N; 1.
 DR PRINTS; PR00148; ENOLASE.
 DR ProDom; PD000902; Enolase; 1.
 DR TIGRFAMs; TIGR01060; eno; 1.
 DR PROSITE; PS00164; ENOLASE; 1.
 DR Lyase; Glycolysis; Magnesium; Multigene family.
 FT ACT SITE 164 164 BY SIMILARITY.
 FT METAL 251 251 MAGNESIUM (BY SIMILARITY).
 FT METAL 302 302 MAGNESIUM (BY SIMILARITY).
 FT METAL 329 329 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 446 AA; 48063 MW; 6266C48914F35198 CRC64;
 Query Match 54.4%; Score 37; DB 1; Length 446;
 Best Local Similarity 63.6%; Pred.No. 85;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PAVVGLSPGEQ 11
 Db 77 PAIVGKDPTQE 87
 |||||
 |||||
 RESULT 14
 NR41 RAT STANDARD; PRT; 597 AA.
 AC P22829;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 10-OCT-2003 (Rel. 42, last annotation update)
 DE Orphan nuclear receptor HMR (Nerve growth factor induced protein I-B)
 DE (NGFI-B) (NUR77).
 GN NR4A1 OR HMR OR NGFI-B.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90166506; PubMed=3272167;
 RA Milbrandt J.;
 RT "Nerve growth factor induces a gene homologous to the glucocorticoid
 RT receptor gene."
 RL Neuron 11:183-188 (1988).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=93361012; PubMed=8395013;
 RA Wilson T.E., Fahrner T.J., Milbrandt J.;
 RT "The orphan receptors NGFI-B and steroidogenic factor 1 establish
 RT monomer binding as a third paradigm of nuclear receptor-DNA
 RT interaction."
 RL Mol. Cell. Biol. 13:5794-5804 (1993).
 RN [3]
 RP DNA BINDING MOTIFS.

RX MEDLINE=92229411; PubMed=1314418;
 RA Wilson T.E., Paulsen R.E., Padgett K.A., Milbrandt J.;
 RT "Participation of non-zinc finger residues in DNA binding by two
 RT nuclear orphan receptors."
 RL Science 256:1107-1110 (1992).
 RN [4]
 RP PHOSPHORYLATION.
 RX MEDLINE=94043340; PubMed=8227042;
 RA Hirata Y., Kiuchi K., Chen H.-C., Milbrandt J., Guroff G.;
 RT "The phosphorylation and DNA binding of the DNA-binding domain of the
 RT orphan nuclear receptor NGFI-B.";
 RL J. Biol. Chem. 268:24808-24812 (1993).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 264-351 IN COMPLEX WITH NBRE,
 RP AND METAL-BINDING.
 RX MEDLINE=99260743; PubMed=10331876;
 RA Meinke G., Sigler P.B.;
 RT "DNA-binding mechanism of the monomeric orphan nuclear receptor
 RT NGFI-B.";
 RL Nat. Struct. Biol. 6:471-477 (1999).
 CC -|- FUNCTION: Probable nuclear receptor. May act concomitantly with
 CC NURR1 in regulating the expression of delayed-early genes during
 CC liver regeneration. Binds the NGFI-B response element (NBRE) 5'-
 CC AAAGGTCA-3'.
 CC -|- SUBUNIT: Binds DNA as a monomer.
 CC -|- SUBCELLULAR LOCATION: Nuclear.
 CC -|- TISSUE SPECIFICITY: Expressed in lung, brain and superior
 CC cervical ganglia. High levels are seen in the adrenal tissue.
 CC -|- INDUCTION: By nerve growth factor and during liver regeneration.
 CC -|- PIM: Phosphorylation of Ser-350 results in decrease in NBRE
 CC binding while phosphorylation of Ser-340 has little effect on it.
 CC -|- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
 CC subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U17254; AAA56770.1; ALT_INIT.
 CC PDB; 1CIT; 26-JUN-00.
 CC TRANSPAC; T00619; --
 CC InterPro; IPR000536; Hormone rec.lig.
 CC InterPro; IPR001723; Steroid receptor.
 CC InterPro; IPR008946; Steroid receptor.
 CC InterPro; IPR001628; Znf_C4steroid.
 CC Pfam; PF00104; hormone_rec; 1.
 CC Pfam; PF00105; zf-C4; 1.
 CC PRINTS; PR00398; STRDHORMONER.
 CC PRINTS; PR00047; STROIDFINGER.
 CC ProDom; PD000035; Znf_C4steroid; 1.
 CC SMART; SM00430; HOL1; 1.
 CC SMART; SM00399; Znf_C4; 1.
 CC PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
 CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 CC Zinc-finger; Phosphorylation; 3D-structure.
 FT DNA_BIND 266 331 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 266 286 C4-TYPE.
 FT ZN_FING 302 326 C4-TYPE.
 FT DOMAIN 408 458 LIGAND-BINDING (POTENTIAL).
 FT DOMAIN 80 91 POLY-SER.
 FT DOMAIN 182 186 POLY-PRO.
 FT DOMAIN 582 585 POLY-PRO.
 FT MOD_RES 340 340 PHOSPHORYLATION (BY PKA).
 FT MOD_RES 350 350 PHOSPHORYLATION (BY PKA).
 FT MUTAGEN 340 340 S->A: LOSS OF PHOSPHORYLATION.
 FT MUTAGEN 350 350 S->A: LOSS OF PHOSPHORYLATION.
 FT MUTAGEN 345 345 R->K: DECREASED NBRE BINDING.
 FT MUTAGEN 348 348 L->V: ALMOST COMPLETE LOSS OF NBRE
 FT BINDING.


```
SQ SEQUENCE 597 AA; 64281 MW; 9CFA987112337E53 CRC64;
Query Match 54.4%; Score 37; DB 1; Length 597;
Best Local Similarity 50.0%; Pred.No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQE 12
| : |||||:
Db 424 PGFELSPGPD 435

RESULT 15
GLMS CLOTE STANDARD; PRT; 607 AA.
AC Q890U2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]
DE (BC 2.6.1.16) (Hexosephosphate aminotransferase) (D-fructose-6-
DE phosphate amidotransferase) (GFAT) (L-glutamine-D-fructose-6-phosphate
DE amidotransferase) (Glucosamine-6-phosphate synthase).
GN GLMS OR C1C02543.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E98;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baumer S., Fricke W.F., Wierer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
CC -!- FUNCTION: Catalyzes the first step in hexosamine metabolism,
CC converting fructose-6P into glucosamine-6P using glutamine as a
CC nitrogen source (By similarity).
CC -!- CATALYTIC ACTIVITY: L-glutamine + D-fructose 6-phosphate = L-
CC glucamate + D-glucosamine 6-phosphate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: In the C-terminal section; belongs to the SIS family.
CC GFAT subfamily.
CC -!- SIMILARITY: Contains 1 type-2 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE015944; AAC37003.1; ALT_INIT.
DR HAMAP; MF 00164; -; 1.
DR InterPro; IPR000583; GATase_2.
DR InterPro; IPR005855; GLMS.
DR InterPro; IPR001347; SIS.
DR Pfam; PF00310; GATase_2; 1.
DR Pfam; PF01380; SIS; 2.
DR TIGRPFAM; TIGR01135; glms; 1.
DR PROSITE; PS00443; GATASE TYPE II; 1.
KW Transferase; Aminotransferase; Glutamine amidotransferase;
KW Complete proteome.
FT INIT MET 0 0 BY SIMILARITY.
FT DOMAIN 1 239 GLUTAMINE AMIDOTRANSFERASE.
FT ACT SITE 1 1 GATASE (BY SIMILARITY).
FT ACT SITE 602 602 ISOMERIZATION FRU-6P (BY SIMILARITY).
SQ SEQUENCE 607 AA; 67694 MW; F4B1CF2EEA37948 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 607;
```

```
Best Local Similarity 61.5%; Pred.No. 1.1e+02;
Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 1 PAVVGLSPGEQE 13
| : |||||:
Db 176 PLIVGL--GEERY 186

Search completed: May 7, 2004, 12:34:33
Job time : 4.68 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:27:20 ; Search time 8.32 Seconds
(without alignments)
150.299 Million cell updates/sec

Title: US-09-786-214a-13

Perfect score: 68

Sequence: 1 PAWGLSPGEQY 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listning first 45 summaries

Database :

PIR 78:*

1: Pirl:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	63.2	211	2 A64490	precorrin-6Y methy
2	43	63.2	623	2 T40991	probable lysophosp
3	43	63.2	821	2 C84304	DNA helicase limpo
4	41	60.3	156	2 F87551	conserved hypotet
5	41	60.3	326	2 T45226	probable N5,N10-me
6	41	60.3	508	2 E70764	probable cobi prote
7	40	58.8	227	2 B90400	hypothetical prote
8	40	58.8	243	2 I54459	MHC H-2K1-k - mous
9	40	58.8	381	2 F75270	cytochrome p450 -
10	40	58.8	540	2 A75250	carboxylesterase,
11	40	58.8	661	2 G84511	hypothetical prote
12	40	58.8	673	2 T50281	probable lysophosp
13	39	57.4	43	2 S21065	Ig kappa chain V r
14	39	57.4	96	2 S45441	Ig kappa chain V r
15	39	57.4	103	2 S19975	Ig kappa chain V r
16	39	57.4	106	2 PS0070	Ig kappa chain V r
17	39	57.4	106	2 PC4282	Ig kappa chain (an
18	39	57.4	107	2 S57444	Ig kappa chain V-J
19	39	57.4	108	2 C30502	Ig kappa chain V r
20	39	57.4	108	2 G44151	Ig kappa chain V r
21	39	57.4	111	2 S23628	Ig kappa chain V r
22	39	57.4	114	2 S54905	Ig kappa chain V r
23	39	57.4	115	1 K3HUVG	Ig kappa chain pre
24	39	57.4	115	1 KVM5L7	Ig kappa chain pre
25	39	57.4	115	2 S11697	Ig kappa chain pre
26	39	57.4	116	2 B25521	Ig kappa chain pre
27	39	57.4	119	2 S41816	Ig kappa chain V r
28	39	57.4	125	2 S40344	Ig kappa chain V-J
29	39	57.4	128	2 PN0445	Ig kappa chain pre

ALIGNMENTS

RESULT 1

A64490
precorrin-6Y methylase homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 18-Aug-2003
C:Accession: A64490
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.; Reich, C.I.; Overbeek, R.; Kirkness, B.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: A64490
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-211 <BUL>
A:Cross-references: GB:U67593; GB:L77117; NID:g2826427; PIDN:AAB99541.1; PID:g1592152; C
C:Genetics:
A:Map position: FOR1500322-1500957
C:Superfamily: precorrin-6Y methylase ChiE

Query Match 63.2%; Score 43; DB 2; Length 211;
Best Local Similarity 54.5%; Pred. No. 6.7;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVGLSPGEQY 13

Db 4 IVGIGPGDREY 14

RESULT 2

T40991
probable lysophospholipase precursor - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T40991
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21962
A:Accession: T40991
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-623 <LYN>
A:Cross-references: EMBL:AL049559; PIDN:CAB40176.1; GSPDB:GNO0068; SPDB:SPCCI450.09C
A:Experimental source: strain 972h-; cosmid c1450
C:Genetics:
A:Gene: SPDB:SPCCI450.09C
A:Map position: 3
C:Superfamily: yeast lysophospholipase
Query Match 63.2%; Score 43; DB 2; Length 623;

Ig kappa chain V-J
Ig kappa chain V r
Ig kappa chain V r
Ig kappa chain - h
Ig kappa chain pre
Ig kappa chain pre
Ig heavy chain V r
Ig heavy chain V-I
uracil permease ho
uracil permease ho
BUD3 protein - yea
Ig heavy chain V r
Ig light chain var
Ig kappa chain V r
probable transcrip
probable transcrip

30 39 57.4 128 2 S40379
31 39 57.4 128 2 A56701
32 39 57.4 129 2 S29627
33 39 57.4 129 2 S40363
34 39 57.4 132 2 S05268
35 39 57.4 144 2 P01006
36 39 57.4 144 2 B30502
37 39 57.4 215 2 A23746
38 39 57.4 428 2 AG1304
39 39 57.4 428 2 AG1676
40 39 57.4 1367 2 S74285
41 38 55.9 86 2 S2649
42 38 55.9 110 2 S60591
43 38 55.9 111 2 PN0537
44 38 55.9 152 2 G75184
45 38 55.9 155 2 G71217

Best Local Similarity 69.2%; Pred. No. 21;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAVVGLSPGQEQY 13
|||
Db 76 PASDGLSTGEQEP 88

RESULT 3
C84304
DNA helicase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84304
R:Ng, W.W.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Laaky, S.;
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jable
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: C84304
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-821 <STO>
A:Cross-references: GB:AE004437; NID:gl0580995; PIDN:AAG19799.1; GSPDB:GN00138
C:Genetics:
A:Gene: hel

Query Match 63.2%; Score 43; DB 2; Length 821;
Best Local Similarity 90.0%; Pred. No. 27;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AVVGLSPGQEQ 11
|||||
Db 326 AVVGLSPAEQ 335

RESULT 4
F87551
conserved hypothetical protein CC2439 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: F87551
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87551
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <STO>
A:Cross-references: GB:AE005673; NID:gl3423984; PIDN:AAK24410.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2439
C:Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

Query Match 60.3%; Score 41; DB 2; Length 156;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 AVVGLSPGQEQ 11
|||||
Db 18 AVVGLDFGEK 27

RESULT 5
T45226
probable N5, N10-methylene-tetrahydromethanopterin reductase (F420-dependent) [imported]
C:Species: Methanobolus tindarius
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000

C;Superfamily: precorrin-6Y methylase CblE

Query Match 58.8%; Score 40; DB 2; Length 227;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 3 VVGLSPGEQY 13
: : : : :
Db 10 IVGVGPDPEY 20

RESULT 8

I54459
MHC H-2K1-k - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I54459
R;Watts, S.; Davis, A.C.; Goodenow, R.S.
Immunogenetics 29, 355-357, 1989
A;Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.
A;Reference number: I54459; MUID:89233303; PMID:2714856
A;Accession: I54459
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-243 <RES>
A;Cross-references: GB:M27134; NID:g199435; PIDN:AAA39610.1; PID:g387456
C;Genetics:
A;Introns: 22/1; 112/1

Query Match 58.8%; Score 40; DB 2; Length 243;
Best Local Similarity 77.8%; Pred. No. 26;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 4 VGLSPGEQ 12

Db 216 LGLSPGEE 224

RESULT 9

F75270
cytochrome P450 - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75270
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75270
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <WHI>
A;Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PID:g646028
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2473
A;Map position: 1

Query Match 58.8%; Score 40; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAVVGLSP 8

Db 53 PAVVGLSP 60

RESULT 10

A75250
carboxylesterase, type B - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: A75250
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: A75250
A;Molecule type: DNA
A;Residues: 1-540 <WHI>
A;Cross-references: GB:AE002092; GB:AE000513; NID:g6460455; PIDN:AAF12163.1; PID:g646045
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2626
A;Map position: 1

C;Superfamily: cholinesterase; cholinesterase homology

Query Match 58.8%; Score 40; DB 2; Length 540;
Best Local Similarity 70.0%; Pred. No. 59;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PAVVGLSPGE 10

Db 512 PQVGLGAPGE 521

RESULT 11

G84511
hypothetical protein At2g13900 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84511
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84511
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-661 <STO>
A;Cross-references: GB:AE002093; NID:g6598598; PIDN:AAF18650.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g13900
A;Map position: 2

Query Match 58.8%; Score 40; DB 2; Length 661;
Best Local Similarity 53.8%; Pred. No. 72;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQY 13

Db 437 PLTKIVPGEQY 449

RESULT 12

T50281
probable lysophospholipase (EC 3.1.1.5) precursor SPAC977.09c [similarity] - fission yeast
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 19-Jan-2001
C;Accession: T50281; F42738
R;Zimmermann, W.; Wambutt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 2000
A;Reference number: Z25053
A;Accession: T50281

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-673 <ZIM>

A;Cross-references: EMBL:AL137130; NID:g6742151; PIDN:CAB69631.1; PID:g6742159; GSPDB:GN
A;Experimental source: strain 972h(-); cosmid c977
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.

Fri May 7 13:19:31 2004

us-09-786-214a-13.rpr

DNA Res. 4, 363-369, 1997

A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:96162722; PMID:9501991
A;Accession: T42738
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 427-673 <YOS>
A;Cross-references: EMBL:D89183; NID:g1749573; PIDN:BAA13845.1; PID:g1749574
A;Experimental source: strain PR745
C;Genetics:
A;Gene: SPDB:SPAC977.09c
A;Map position: 1
A;Introns: 651/3

C;Description: catalyzes the hydrolysis of 2-lysophosphatidylcholine to glycerophosphocholine
A;Superfamily: yeast lysophospholipase
C;Keywords: carboxylic ester hydrolase

Query Match 58.8%; Score 40; DB 2; Length 673;
Best Local Similarity 61.5%; Pred. No. 74;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAVVGLSPGGEY 13
||| : ||| :
Db 83 PASEGLNEGEQSY 95

RESULT 13

S21065
Ig kappa chain V region (anti-RH(D)) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 24-May-1996 #text_change 09-May-1997
C;Accession: S21065
R;Dlouha, A.; Lecroisey, A.; Henschen, A.; Rouger, P.; Keil, B.
Protein Seq. Data Anal. 4, 317-318, 1991
A;Title: Subgroup assignment of a human monoclonal anti-Rh(D) antibody.
A;Reference number: S21065; MUID:92253544; PMID:1812483
A;Accession: S21065
A;Molecule type: protein
A;Residues: 1-43 <DLO>
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 57.4%; Score 39; DB 2; Length 43;
Best Local Similarity 63.6%; Pred. No. 6.3;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVVGLSPGGEQ 11
||| : ||| :
Db 8 PATLSLSPGER 18

RESULT 14

S45441
Ig kappa chain V region - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jan-2000
C;Accession: S45441
R;Cox, J.P.L.; Tomlinson, I.M.; Winter, G.
Eur. J. Immunol. 24, 827-836, 1994
A;Title: A directory of human germ-line V(kappa) segments reveals a strong bias in their
A;Reference number: S45324; MUID:94200218; PMID:8149953
A;Accession: S45441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-96 <COX>
A;Cross-references: EMBL:Z27500
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;16-91/Domain: immunoglobulin homology <IMM>

Query Match 57.4%; Score 39; DB 2; Length 96;
Best Local Similarity 63.6%; Pred. No. 15;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVVGLSPGGEQ 11
||| : ||| :
Db 8 PATLSLSPGER 18

RESULT 15

S19975
Ig kappa chain V region (M-T408) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 21-Jan-2000
C;Accession: S19975
R;Weissenborn, W.; Riethmuller, G.; Weiss, E.M.; Rieber, E.P.
submitted to the EMBL Data Library, March 1992
A;Description: Structural characterization of CD4 mAb.
A;Reference number: S19963
A;Accession: S19975
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-103 <WEI>
A;Cross-references: EMBL:X65097; NID:g52296; PIDN:CAA46225.1; PID:g52297
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin
F;11-85/Domain: immunoglobulin homology <IMM>

Query Match 57.4%; Score 39; DB 2; Length 103;
Best Local Similarity 63.6%; Pred. No. 16;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVVGLSPGGEQ 11
||| : ||| :
Db 3 PATLSLSPGER 13

Search completed: May 7, 2004, 12:39:07
Job time : 9.48667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:27:20 ; Search time 9.6 Seconds
(without alignments)
150.299 Million cell updates/sec

Title: US-09-786-214A-15

Perfect score: 75
Sequence: 1 AGLPAVVGLSPGEQE 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: *
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	66.7	381	2	Cytochrome P450 -
2	43	57.3	659	1	Gene 18 protein -
3	43	57.3	659	1	tail sheath protei
4	43	57.3	821	2	DNA helicase [impo
5	42	56.0	326	2	probable N5.N10-me
6	42	56.0	508	2	probable cobl prot
7	41	54.7	156	2	conserved hypotet
8	41	54.7	164	2	ISCC1, transposase
9	41	54.7	177	2	hypothetical prote
10	41	54.7	214	2	hypothetical prote
11	40	53.3	130	2	hypothetical prote
12	40	53.3	243	2	MHC H-2K1-k - mous
13	40	53.3	358	1	Aut protein - toma
14	40	53.3	359	2	fructose-bisphosph
15	40	53.3	359	2	uracil permease ho
16	40	53.3	428	2	uracil permease ho
17	40	53.3	428	2	carboxylesterase,
18	40	53.3	540	2	hypothetical prote
19	40	53.3	563	2	probable lysophosp
20	40	53.3	623	2	yhgF protein - Esc
21	40	53.3	740	1	hypothetical prote
22	40	53.3	740	2	hypothetical prote
23	40	53.3	740	2	hypothetical prote
24	40	53.3	776	2	probable transcrip
25	40	53.3	1340	2	phosphoglycan core
26	40	53.3	1506	2	phosphonositide 3
27	40	53.3	2327	2	aggreacan - bovine
28	39	52.0	43	2	Ig kappa chain V r
29	39	52.0	96	2	Ig kappa chain V r

ALIGNMENTS

RESULT 1

F75270
Cytochrome P450 - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C/Accession: F75270
R/White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; M. Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036896; PMID:10567266
A/Accession: F75270
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-381 <WHI>
A/Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PTD:g64602
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR2473
A/Map position: 1

30	39	52.0	103	2	S19975	Ig kappa chain V r
31	39	52.0	106	2	P90070	Ig kappa chain V r
32	39	52.0	106	2	PC4282	Ig kappa chain (an
33	39	52.0	107	2	S57444	Ig kappa chain V-J
34	39	52.0	108	2	C30502	Ig kappa chain V r
35	39	52.0	108	2	S33988	Ig kappa chain V r
36	39	52.0	108	2	G44151	Ig kappa chain V r
37	39	52.0	111	2	S23628	Ig kappa chain V r
38	39	52.0	114	2	S54905	Ig kappa chain V r
39	39	52.0	115	1	K3HUVG	Ig kappa chain V r
40	39	52.0	115	1	KVMSL7	Ig kappa chain pre
41	39	52.0	115	2	S11657	Ig kappa chain pre
42	39	52.0	115	2	G72587	hypothetical prote
43	39	52.0	116	2	B25521	Ig kappa chain pre
44	39	52.0	119	2	S41816	Ig kappa chain V r
45	39	52.0	125	2	S40344	Ig kappa chain V-J

Query Match 66.7%; Score 50; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2	GLPAVVGLSP 11
Db	51	GLPAVVGLSP 60

RESULT 2

GKBPT4
gene 18 protein - phage T4
N/Alternate names: tail sheath protein gp18
C/Species: phage T4
A/Note: host Escherichia coli
C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 23-Jul-1999
C/Accession: JF0021
R/Arisaka, F.; Nakako, T.; Takahashi, H.; Ishii, S.
J. Virol. 62, 1186-1193, 1988
A/Title: Nucleotide sequence of the tail sheath gene of bacteriophage T4 and amino acid
A/Reference number: JF0021; MUID:88155753; PMID:2964531
A/Accession: JF0021
A/Molecule type: DNA
A/Residues: 1-659 <ARI>
A/Cross-references: GB:M19085; EMBL:M36959; NID:g215949; PIDN:AAA32541.1; PTD:g215950
C/Genetics:
A/Gene: 18
A/Map position: 97 352-99 328
C/Superfamily: phage T4 gene 18 protein

Query Match 57.3%; Score 43; DB 1; Length 659;
 Best Local Similarity 66.7%; Pred. No. 56;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPAVVLSPGE 13
 | : | | | | |
 DB 218 GIPGVVLYPGE 229

RESULT 3
 A:Accession: A60885
 C:Species: tail sheath protein - phage T4
 C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 03-May-1996
 C:Accession: A60885
 R:Arisaka, F.; Nakako, T.; Kumazaki, T.; Ishii, S.
 J. Protein Chem. 6, 245-251, 1987
 A:Title: Primary structure of the tail sheath protein of bacteriophage T4 and its gene.
 A:Reference number: A60885
 A:Accession: A60885
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-659 <ARI>
 C:Superfamily: phage T4 gene 18 protein

Query Match 57.3%; Score 43; DB 2; Length 659;
 Best Local Similarity 66.7%; Pred. No. 56;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPAVVLSPGE 13
 | : | | | | |
 DB 218 GIPGVVLYPGE 229

RESULT 4
 A:Accession: C84304
 C:Species: DNA helicase [imported] - Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: C84304
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
 A:Title: Genome sequence of Halobacterium species NRC-1
 A:Reference number: A84160; MUID:20504483; PMID:11016950
 A:Accession: C84304
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-821 <STO>
 A:Cross-references: GB:AE004437; NID:g10580995; PIDN:AAG19799.1; GSPDB:GN00138
 C:Genetics: hel
 A:Gene: hel

Query Match 57.3%; Score 43; DB 2; Length 821;
 Best Local Similarity 90.0%; Pred. No. 70;
 Matches 9; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

QY 5 AVVGLSPGEQ 14
 | : | | | | |
 DB 326 AVVGLSPAEQ 335

RESULT 5
 T45226
 C:Species: Probable N5,N10-methylene-tetrahydromethanopterin reductase (F420-dependent) [imported]
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
 C:Accession: T45226
 R:Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G.;
 submitted to the EMBL Data Library, September 1998
 A:Description: The F420H2-dehydrogenase from Methanobolus tindarius: Cloning of the ffd

A:Reference number: Z22947
 A:Accession: T45226
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-326 <WES>
 A:Cross-references: EMBL:AJ011519; PIDN:CAB56639.1
 A:Experimental source: DSM 2278
 C:Genetics: ffd
 A:Gene: ffd

Query Match 56.0%; Score 42; DB 2; Length 326;
 Best Local Similarity 57.1%; Pred. No. 40;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLPAVVLSPGEQ 14
 | : | : | | | | |
 DB 84 SGGRAILLGLSPGEQ 97

RESULT 6
 E70764
 C:Species: probable cobi protein - Mycobacterium tuberculosis (strain H37RV)
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: E70764
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: E70764
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-508 <COL>
 A:Cross-references: GB:273966; GB:AL123456; NID:g3261577; PIDN:CAA98214.1; PID:e246996;
 A:Experimental source: strain H37RV
 C:Genetics: cobi
 A:Gene: cobi

Query Match 56.0%; Score 42; DB 2; Length 508;
 Best Local Similarity 57.1%; Pred. No. 62;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 GLPAVVLSPGEQ 15
 | : | | | | | : |
 DB 247 GTVAVVGLSPGDS 260

RESULT 7
 F87551
 C:Species: conserved hypothetical protein CC2439 [imported] - Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: F87551
 R:Nierman, W.C.; Feidblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.D.; Haft, D.H.; Kolo
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: F87551
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-156 <STO>
 A:Cross-references: GB:AE005673; NID:g13423984; PIDN:AAK24410.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC2439
 C:Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

Query Match 54.7%; Score 41; DB 2; Length 156;
 Best Local Similarity 80.0%; Pred. No. 28;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 AVVGLSPGEQ 14
|||||
Db 18 AVVGLDPGEK 27

RESULT 8

A87399
ISC1, transposase OrfB [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 18-Jul-2001
C:Accession: A87399; E87402; E87502; E87570
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapero, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87399
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-164 <STO>
A:CROSS-references: GB:AE005673; NID:G13422533; PIDN:AAK23189.1; GSPDB:GN00148
A:Accession: E87402
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-164 <ST2>
A:CROSS-references: GB:AE005673; NID:G13422563; PIDN:AAK23317.1; GSPDB:GN00148
A:Accession: E87502
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-164 <ST3>
A:CROSS-references: GB:AE005673; NID:G13423519; PIDN:AAK24017.1; GSPDB:GN00148
A:Accession: E87570
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-164 <ST4>
A:CROSS-references: GB:AE005673; NID:G13424163; PIDN:AAK24561.1; GSPDB:GN00148
C:Genetics:
C:Superfamily: Streptomyces coelicolor probable transposase SC609.35c

Query Match 54.7%; Score 41; DB 2; Length 164;
Best Local Similarity 53.8%; Pred. No. 29;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 GLPAVVGSLPGEQ 14
|||||
Db 49 GLPVRIALGGEQ 61

RESULT 9

B72580
hypothetical protein APE1922 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: B72580
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: B72580
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <KAW>
A:CROSS-references: DDBJ:AP000062; NID:G5105244; PIDN:BAA80927.1; PID:d1044713; PID:G510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1922

Query Match 54.7%; Score 41; DB 2; Length 177;

Best Local Similarity 53.8%; Pred. No. 31;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLPAVVGSLPGEQ 14
|||||
Db 151 GVPVVLGEPWDR 163

RESULT 10

A83416
hypothetical protein PA1825 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83416
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.K.; Kas, A.; Larbig, K.; Lim
; Lozy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic path
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83416
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-214 <STO>
A:CROSS-references: GB:AE004609; GB:AE004091; NID:G9947810; PIDN:AAG05214.1; GSPDB:GN00
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA1825

Query Match 54.7%; Score 41; DB 2; Length 214;
Best Local Similarity 63.6%; Pred. No. 38;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLPAVVGSLPGEQ 12
|||||
Db 113 GLAALIGLAPG 123

RESULT 11

D72783
hypothetical protein APE0253 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: D72783
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: D72783
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-130 <KAW>
A:CROSS-references: DDBJ:AP000058; NID:G5103388; PIDN:BAA79166.1; PID:d1042942; PID:G51
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0253
C:Superfamily: Aeropyrum pernix hypothetical protein APE0253

Query Match 53.3%; Score 40; DB 2; Length 130;
Best Local Similarity 58.3%; Pred. No. 33;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPAVVGSLPGEQ 13
|||||
Db 5 GVPVVLGEPGE 16

RESULT 12

I54459
MHC H-2K1-k - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999

C:Accession: I54459
 R:Waters, S.; Davis, A.C.; Goodenow, R.S.
 Immunogenetics 29, 355-357, 1989
 A:Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.
 A:Reference number: I54459; MUID:89233303; PMID:2714856
 A:Accession: I54459
 A:Status: preliminary; translated from GB/EMBL/DBU
 A:Molecule type: DNA
 A:Residues: 1-243 <RES>
 A:Cross-references: GB:M27134; NID:G199435; PIDN:AAA39610.1; PID:G387456
 C:Genetics:
 A:Introns: 22/1; 112/1

Query Match 53.3%; Score 40; DB 2; Length 243;
 Best Local Similarity 77.8%; Pred. No. 62;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 7 VGLSPGEQE 15
 :|||:|:
 Db 216 LGLSPGEEE 224

RESULT 13
 JQ1870
 A11 protein - tomato mottle virus (isolate Florida)
 C:Species: tomato mottle virus
 C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 07-May-1999
 C:Accession: JQ1870
 R:Abouzid, A.M.; Polston, J.E.; Hiebert, E.
 J. Gen. Virol. 73, 3225-3229, 1992
 A:Title: The nucleotide sequence of tomato mottle virus, a new geminivirus isolated from
 A:Reference number: JQ1869; MUID:93107858; PMID:1469361
 A:Accession: JQ1870
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-358 <ABO>
 A:Cross-references: GB:L14460
 C:Genetics:
 A:Map position: segment A
 C:Superfamily: tomato golden mosaic virus A11 protein

Query Match 53.3%; Score 40; DB 1; Length 358;
 Best Local Similarity 58.3%; Pred. No. 92;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLPAVVGLSPE 13
 :|||:|:
 Db 296 GIPAVLCNPGE 307

RESULT 14
 S58167
 fructose-bisphosphate aldolase (EC 4.1.2.13) - garden pea
 N:Alternate names: fructose-1,6-bisphosphate aldolase
 C:Species: Pisum sativum (garden pea)
 C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
 C:Accession: S58167
 R:Pelzer-Reith, B.; Schnarrenberger, C.
 submitted to the EMBL Data Library, July 1995
 A:Description: Characterization of cDNA clones and expression of two cytosolic fructose-
 A:Reference number: S58167
 A:Accession: S58167
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-359 <PEL>
 A:Cross-references: EMBL:X89829; NID:G927504; PIDN:CAA61947.1; PID:G927505
 C:Superfamily: fructose-bisphosphate aldolase
 C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 53.3%; Score 40; DB 2; Length 359;
 Best Local Similarity 60.0%; Pred. No. 92;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLPAVVGLSPEQE 15
 :|||:|:
 Db 258 AAVPAVVGSLGGQSE 272

RESULT 15
 T48396
 fructose-bisphosphate aldolase-like protein - Arabidopsis thaliana
 N:Alternate names: protein F17C15.110
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
 C:Accession: T48396
 R:Revan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.
 submitted to the Protein Sequence Database, March 2000
 A:Reference number: Z24492
 A:Accession: T48396
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-359 <BEV>
 A:Cross-references: EMBL:AL162506
 A:Experimental source: cultivar Columbia; BAC clone F17C15
 C:Genetics:
 A:Map position: 5
 A:Introns: 10/1; 86/1
 A:Note: F17C15.110
 C:Superfamily: fructose-bisphosphate aldolase

Query Match 53.3%; Score 40; DB 2; Length 359;
 Best Local Similarity 60.0%; Pred. No. 92;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLPAVVGSLPEQE 15
 :|||:|:
 Db 258 AAVPAVVGSLGGQSE 272

Search completed: May 7, 2004, 12:39:07
 Job time : 9.76667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 4.68 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-14
Perfect score: 65
Sequence: 1 LPAVVGLSPGGEQ 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	63.1	156	1	RVUX_CAUCR
2	41	63.1	326	1	MER_METTI
3	40	61.5	633	1	PLB5_SCHPO
4	39	60.0	115	1	KV3I_HUMAN
5	39	60.0	115	1	KV5I_MOUSE
6	39	60.0	508	1	COBI_MYCTU
7	39	60.0	597	1	NR41_RAT
8	39	60.0	1402	1	N160_MOUSE
9	39	60.0	1636	1	BUD3_YEAST
10	38	58.5	390	1	COBL_MYCTU
11	38	58.5	429	1	RS1_LEULA
12	38	58.5	507	1	CAT4_PICAN
13	38	58.5	516	1	C4AD_DROME
14	38	58.5	699	1	EFG_HAEIN
15	38	58.5	700	1	EFG_PASMU
16	37.5	57.7	827	1	M4K1_MOUSE
17	37	56.9	280	1	PMXA_MOUSE
18	37	56.9	281	1	PMXA_RAT
19	37	56.9	446	1	COBJ_ARCFU
20	37	56.9	446	1	ENOI_MAIZE
21	37	56.9	557	1	PUR6_VIGAC
22	37	56.9	637	1	MUTL_CAUCR
23	37	56.9	658	1	VG18_BP74
24	37	56.9	813	1	CADM_MOUSE
25	37	56.9	985	1	4ET_HUMAN
26	37	56.9	2269	1	RRKL_SV41
27	36.5	56.2	374	1	RGPS_BOVIN
28	36	55.4	129	1	KV3H_HUMAN
29	36	55.4	211	1	COBL_METJA
30	36	55.4	279	1	TYSY_CAUCR
31	36	55.4	342	1	HOPK_AZOVI
32	36	55.4	348	1	HOPV_AZOVI
33	36	55.4	359	1	ALF2_PEA

34	36	55.4	359	1	ALF_CICAR	O65735 cicer ariet
35	36	55.4	590	1	MUTL_THETIN	O8ra70 thermoanaer
36	36	55.4	598	1	NR41_CANFA	P51666 canis famil
37	36	55.4	750	1	ELS_CHICK	P07916 gallus gall
38	35.5	54.6	833	1	M4K1_HUMAN	O92918 homo sapien
39	35	53.8	100	1	KV3C_HUMAN	P01621 homo sapien
40	35	53.8	109	1	KV3B_HUMAN	P01620 homo sapien
41	35	53.8	109	1	KV3D_HUMAN	P01622 homo sapien
42	35	53.8	109	1	KV3E_HUMAN	P01623 homo sapien
43	35	53.8	109	1	KV3G_HUMAN	P04206 homo sapien
44	35	53.8	129	1	KV3L_HUMAN	P18135 homo sapien
45	35	53.8	129	1	KV3M_HUMAN	P18136 homo sapien

ALIGNMENTS

RESULT 1
ID RVUX_CAUCR STANDARD; PRT; 156 AA.
AC Q9A5K8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative Holliday Junction resolvase (EC 3.1.1.-).
GN CC2439.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_taxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
Petrocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
Kolony J.F., Smit J., Duran M.B., Khouri H., Shetty J., Berry K.,
Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
"Complete genome sequence of Caulobacter crescentus.";
"-!- FUNCTION: Could be a nuclease that resolves Holliday junction
intermediates in genetic recombination.
"-!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
"-!- SIMILARITY: Belongs to the YGF HJR family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
or send an email to license@ebi.ac.uk).

EMBL; AB005913; AAK24410.1; -;
DR PIR; F87551; F87551.
DR TIGR; CC2439; -;
HAMAP; MF_00651; -; 1.
DR InterPro; IPR005227; Cons_hypoth250.
DR InterPro; IPR006641; YqgFc.
DR Pfam; PF03652; UPF0081; 1.
DR SMART; SM00732; YqgFc; 1.
DR TIGRFAMs; TIGR00250; TIGR00250; 1.
KW Hydrolase; Nuclease; DNA repair; DNA recombination; Complete proteome.
SQ SEQUENCE 156 AA; 17142 MW; 21F54D8648396141 CRC64;

Query Match 63.1%; Score 41; DB 1; Length 156;
Best Local Similarity 80.0%; Pred. No. 6.8;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 AVVGLSPGGEQ 12

```

Db      18 AVVGLDPGEK 27
||||| |||:
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
NCBI_TaxID=4896;
[1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=972;
RC      MEDLINE=2184901; PubMed=11859360;
RA      Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA      Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA      Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA      Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA      Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA      Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA      James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA      Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA      Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA      Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA      Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA      Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA      Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA      Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA      Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA      Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA      Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA      Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA      Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA      Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA      Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA      Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA      Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA      Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT      "The genome sequence of Schizosaccharomyces pombe.";
RT      Nature 415:871-880(2002).
RL      PATHWAY: Methanogenesis from carbon dioxide; fifth step.
CC      -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC      -!- SIMILARITY: Belongs to the mer family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AJ011519; CAB56639.1; -.
CC      F1R; T45226; T45226.
CC      HAMAP; MF 01091; 1.
CC      InterPro; IPR002103; Bac_luciferase.
CC      Pfam; PF00296; bac_luciferase; 1.
CC      Methanogenesis; One-carbon metabolism; Oxidoreductase.
CC      SQ      SEQUENCE 326 AA; 34043 MW; 16F3AB9733A45D82 CRC64;
Query Match 63.1%; Score 41; DB 1; Length 326;
Best Local Similarity 70.0%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY      3 AVVGLSPGEQ 12
|:|:| |||
Db      88 AILGLGFEQ 97
||||| |||:
RESULT 3
PLB5 SCHPO STANDARD; PRT; 633 AA.
AC      Q9Y7N6;
AT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DE      Putative lysophospholipase C1450.09c precursor (EC 3.1.1.5)
DE      (Phospholipase B).
DE      SPC1450.09c.
OS      Schizosaccharomyces pombe (Fission yeast).
OC      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

```

FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 633 AA; 68292 MW; 49871B2955893D19 CRC64;

Query Match 61.5%; Score 40; DB 1; Length 633;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 13
||| ||| ||| |||
DB 76 PASDGLSTGEQ 87

RESULT 4
KV31_HUMAN STANDARD; PRT; 115 AA.
AC P04433;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig kappa chain V-III region Vg precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed=6440122;
RX MEDLINE=85087932; PubMed=6440122;
RA Pech M., Zachau H.G.;
RT "Immunoglobulin genes of different subgroups are interdigitated
within the VK locus.";
RL Nucleic Acids Res. 12:9229-9236(1984).
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; X01668; -; NOT ANNOTATED_CDS.
PIR; A01900; K3HUVG.
HSP; P80362; 1WTL.
GO; GO:0005576; C:extracellular; NAS.
GO; GO:0003823; F:antigen binding; NAS.
GO; GO:0006955; F:immune response; NAS.
InterPro; IPR007110; Ig-like.
Pfam; PF00047; Ig_v.
SMART; SM00406; IGV; 1.
PROSITE; PS00835; IG_LIKE; 1.
Immunoglobulin V region; Signal.
FT SIGNAL 1 20
FT CHAIN 21 >115 IG KAPPA CHAIN V-III REGION VG.
FT DOMAIN 21 43 FRAMEWORK-1.
FT DOMAIN 44 54 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 55 69 FRAMEWORK-2.
FT DOMAIN 70 76 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 77 108 FRAMEWORK-3.
FT DOMAIN 109 115 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 43 108 BY SIMILARITY.
FT NON TER 115 115
SQ SEQUENCE 115 AA; 12575 MW; 2DE47CDA3A17D555 CRC64;

Query Match 60.0%; Score 39; DB 1; Length 115;
Best Local Similarity 63.6%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
||| : |||||
DB 28 PATLSVSPGER 38

RESULT 5
KV51_MOUSE STANDARD; PRT; 115 AA.
AC P01642;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig kappa chain V-V region L7 precursor (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. PubMed=6264318;
RX MEDLINE=81220975; PubMed=6264318;
RA Pech M., Hochli J., Schnell H., Zachau H.G.;
RT "Differences between germ-line and rearranged immunoglobulin V kappa
coding sequences suggest a localized mutation mechanism.";
RL Nature 291:668-670(1981).
CC -!- MISCELLANEOUS: THERE APPEAR TO BE TWO POSSIBLE SPICE JUNCTIONS AT
THE 3' END OF THE INTRON. THE ALTERNATE WOULD CODE FOR A PROTEIN
LACKING RESIDUES 17-19.
PIR; A01925; KVMSL7.
PDB; 1J10; 18-FEB-03.
PDB; 1J1P; 18-FEB-03.
PDB; 1J1X; 18-FEB-03.
InterPro; IPR007110; Ig-like.
InterPro; IPR003596; Ig_v.
Pfam; PF00047; Ig; 1.
SMART; SM00406; IGV; 1.
PROSITE; PS00835; IG_LIKE; 1.
Immunoglobulin V region; Signal; 3D-structure.
FT SIGNAL 1 20
FT CHAIN 21 >115 IG KAPPA CHAIN V-V REGION L7.
FT DOMAIN 21 43 FRAMEWORK-1.
FT DOMAIN 44 54 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 55 69 FRAMEWORK-2.
FT DOMAIN 70 76 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 77 108 FRAMEWORK-3.
FT DOMAIN 109 >115 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 43 108 BY SIMILARITY.
FT NON TER 115 115
SQ SEQUENCE 115 AA; 12615 MW; C17BEC758C577E00 CRC64;

Query Match 60.0%; Score 39; DB 1; Length 115;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
||| : |||||
DB 28 PATLSVSPGER 38

RESULT 6
COBI_MYCTU STANDARD; PRT; 508 AA.
ID COBI_MYCTU
AC Q10677;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cobalamin biosynthesis protein COBIJ [includes: Precorrin-2 C20-
DE methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine-precorrin-2
DE methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.1.-)].
GN COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.

RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98235987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Braham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=2226494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Biswal W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC PRECORRIN-3A (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC adenosyl-L-homocysteine + precorrin-3A.
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC AND CBIL/CBOL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; Z73966; CAA98214.1; -;
DR EMBL; AE007063; AAK46406.1; -;
DR EMBL; BX248341; CAD96945.1; -;
DR F1R; E70764; E70764.
DR TIGR; MT2126; -;
DR TubercuList; RV2066; -;
DR InterPro; IPR006364; Cobi CbiL.
DR InterPro; IPR006363; CobiJ.
DR InterPro; IPR000878; Cor/por Metransf.
DR Pfam; PF00590; TP methylase; 2.
DR TIGRFAMs; TIGR01467; cobi cbiL; 1.
DR TIGRFAMs; TIGR01466; cobi cbiH; 1.
DR PROSITE; PS00839; SUMT_1; 1.
DR PROSITE; PS00840; SUMT_2; 1.
DR Cobalamin biosynthesis; Porphyryrin biosynthesis; Transfexase;
KW Methyltransferase; Multifunctional enzyme; Complete proteome;
FT DOMAIN 1 243 PRECORRIN-2 C20-METHYLTRANSFERASE.
FT DOMAIN 244 508 PRECORRIN-3 METHYLASE.
SQ SEQUENCE 508 AA; 53910 MW; 95AC066F022C4DC1 CRC64;

Query Match

60.0%; Score 39; DB 1; Length 508;

Best Local Similarity 63.6%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 3 AVUGLSPGEQE 13
DB 250 AVUGLSPGDS 260

RESULT 7
NR41 RAT
ID NR41 RAT STANDARD; PRT; 597 AA.
AC F22829;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Orphan nuclear receptor HMR (Nerve growth factor induced protein I-B)
DE (NGFI-B) (NUR77).
GN NR4A1 OR HMR OR NGFI-B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]_TaxID=10116;
RP SEQUENCE FROM N.A.
RX MEDLINE=90166506; PubMed=3272167;
RA Milbrandt J.;
RT "Nerve growth factor induces a gene homologous to the glucocorticoid
RT receptor gene.";
RL Neuron 1:183-188 (1988).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=93361012; PubMed=8395013;
RA Wilson T.E., Fahrner T.J., Milbrandt J.;
RT "The orphan receptors NGFI-B and steroidogenic factor 1 establish
RT monomer binding as a third paradigm of nuclear receptor-DNA
RT interaction.";
RL Mol. Cell. Biol. 13:5794-5804 (1993).
RN [3]
RP DNA BINDING MOTIFS.
RX MEDLINE=92229411; PubMed=1314418;
RA Wilson T.E., Paulsen R.E., Padgett K.A., Milbrandt J.;
RT "Participation of non-zinc finger residues in DNA binding by two
RT nuclear orphan receptors.";
RL Science 256:107-110 (1992).
RN [4]
RP PHOSPHORYLATION.
RX MEDLINE=94043340; PubMed=8227042;
RA Hirata Y., Kuchi K., Chen H.-C., Milbrandt J., Guroff G.;
RT "The phosphorylation and DNA binding of the DNA-binding domain of the
RT orphan nuclear receptor NGFI-B.";
RL J. Biol. Chem. 268:24808-24812 (1993).
RN [5]
RX X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 264-351 IN COMPLEX WITH NRE,
RP AND METAL-BINDING.
RX MEDLINE=99260743; PubMed=10331876;
RA Meinke G., Sigler P.B.;
RT "DNA-binding mechanism of the monomeric orphan nuclear receptor
RT NGFI-B.";
RL Nat. Struct. Biol. 6:471-477 (1999).
CC -!- FUNCTION: Probable nuclear receptor. May act concomitantly with
CC NUR1 in regulating the expression of delayed-early genes during
CC liver regeneration. Binds the NGFI-B response element (NBRE) 5'-
CC AAAAGGTCA-3'.
CC -!- SUBUNIT: Binds DNA as a monomer.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Expressed in lung, brain and superior
CC cervical ganglia. High levels are seen in the adrenal tissue.
CC -!- INDUCTION: By nerve growth factor and during liver regeneration.
CC -!- PM: Phosphorylation of Ser-350 results in decrease in NBRE
CC binding while phosphorylation of Ser-340 has little effect on it.
CC -!- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
CC subfamily.


```

CC      METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; Z73966; CA398225.1; -.
CC      EMBL; AEC07063; AAK46412.1; -.
CC      PIR; C70765; C70765.
CC      TIGR; MT2132; -.
CC      Tuberculin; Rv2072c; -.
CC      InterPro; IPR006365; COBL.
CC      InterPro; IPR000878; Cor/por Metransf.
CC      InterPro; IPR000051; SAM bind.
CC      Pfam; PF00590; TP methylase; 1.
CC      TIGRPFAMs; TIGR01468; cobL cbiET; 1.
CC      Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW      Methyltransferase; Complete proteome.
FT      CONFLICT 205 205 L -> P (IN REF. 2).
FT      CONFLICT 327 327 D -> H (IN REF. 2).
SQ      SEQUENCE 390 AA; 41854 MW; FB42EPF7562F21F3 CRC64;

Query Match      58.5%; Score 38; DB 1; Length 390;
Best Local Similarity 88.9%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 LPAVUGLSP 9
DB      55 LPAVUGLSP 63

RESULT 11
RS1_LEULA
ID      RS1_LEULA STANDARD; PRT; 429 AA.
AC      P50889; P71450;
DT      01-OCT-1996 (Rel. 34, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      40S ribosomal protein S1.
GN      RPS1
OS      Leuconostoc lactis.
OC      Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX      NCBI_TaxID=1246;
FN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97186703; PubMed=9034319;
RA      Yamit-Hezi A., Levy Z., Neuman S., Nudel U.;
RT      "A Leuconostoc lactis protein with homology to ribosomal protein S1
RT      shares common epitopes and common DNA binding properties with a
RT      mammalian DNA binding nuclear factor.";
RL      Gene 185:99-103(1997).
RL      [2]
RP      SEQUENCE OF 24-429 FROM N.A.
RX      MEDLINE=95237615; PubMed=7721096;
RA      Eklund E.A., Lee S.W., Skalkin D.G.;
RT      "Cloning of a cDNA encoding a human DNA-binding protein similar to
RT      ribosomal protein S1.";
RL      Gene 155:231-235(1995).
RL      [3]
RP      SEQUENCE OF 78-429 FROM N.A.
RX      MEDLINE=96164600; PubMed=8568274;
RA      Tsuzaka K., Leu A.K., Frank M.B., Movafagh B.F., Koscec M.,
RA      Winkler T.H., Kalden J.R., Reichlin M.;
RT      "Lupus autoantibodies to double-stranded DNA cross-react with
RT      ribosomal protein S1.";
RL      J. Immunol. 156:1668-1675(1996).
CC      -!- FUNCTION: EXHIBITS PREFERENTIAL BINDING TO SINGLE-STRANDED AND
CC      DOUBLE-STRANDED DNA AND A LOW BINDING AFFINITY FOR RNA.
CC      -!- SIMILARITY: Belongs to the S1P family of ribosomal proteins.

```

```

CC      -!- SIMILARITY: Contains 4 S1 motif domains.
CC      -!- CAUTION: WAS ORIGINALLY (REF.2 AND REF.3) THOUGHT TO ORIGINATE
CC      FROM HUMAN BUT IS MOST PROBABLY THE RESULT OF A CDNA LIBRARY
CC      CONTAMINATION BY L.LACTIS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; U24086; AAB08978.1; -.
CC      EMBL; U05589; AAA77669.1; -.
CC      EMBL; U27517; AAA97575.1; -.
CC      HSSP; P05055; ISRO.
CC      InterPro; IPR008994; Nucleic acid OB.
CC      InterPro; IPR000110; Ribosomal_S1.
CC      InterPro; IPR003029; S1.
CC      Pfam; PF00575; S1; 4.
CC      PRINTS; PR00681; RIBOSOMALS1.
CC      SMART; SM00316; S1; 4.
CC      PROSITE; PS0126; S1; 4.
KW      Ribosomal protein; Repeat; RNA-binding.
FT      DOMAIN 55 128 S1 MOTIF 1.
FT      DOMAIN 144 211 S1 MOTIF 2.
FT      DOMAIN 231 299 S1 MOTIF 3.
FT      DOMAIN 316 385 S1 MOTIF 4.
FT      CONFLICT 24 24 S -> G (IN REF. 2).
FT      CONFLICT 122 122 A -> S (IN REF. 3).
FT      CONFLICT 217 217 L -> R (IN REF. 2 AND 3).
SQ      SEQUENCE 429 AA; 46386 MW; 92AC82605F39D9DFC CRC64;

Query Match      58.5%; Score 38; DB 1; Length 429;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 AVVGLSPGEQ 12
DB      71 AVVGLSTGEE 80

RESULT 12
CATA_PICAN
ID      CATA_PICAN STANDARD; PRT; 507 AA.
AC      P30263;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Peroxisomal catalase (BC 1.11.1.6).
GN      PXP9 OR PXP-9.
OS      Pichia angusta (Yeast) (Hansenula polymorpha).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Pichia.
OX      NCBI_TaxID=4905;
FN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=ATCC 34438;
RX      MEDLINE=92299073; PubMed=1607006;
RA      Didion T., Roggenkamp R.O.;
RT      "Targeting signal of the peroxisomal catalase in the methylotrophic
RT      yeast Hansenula polymorpha.";
RL      FEBS Lett. 303:113-116(1992).
CC      -!- FUNCTION: Occurs in almost all aerobically respiring organisms and
CC      serves to protect cells from the toxic effects of hydrogen
CC      peroxide.
CC      -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC      -!- COFACTOR: Heme group.
CC      -!- SUBUNIT: Homotetramer.
CC      -!- SUBCELLULAR LOCATION: Peroxisomal.
CC      -!- SIMILARITY: Belongs to the catalase family.

```


This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X56501; CAA39856.1; --
 DR PIR; S23422; S23422.
 DR HSSP; P15202; IA4E.
 DR InterPro; IPR002226; Catalase.
 DR Pfam; PF00199; catalase; 1.
 DR PRINTS; PR00067; CATALASE.
 DR ProDom; PD000510; Catalase; 1.
 DR PROSITE; PS00437; CATALASE_1; 1.
 DR PROSITE; PS00438; CATALASE_2; 1.
 DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
 KW Peroxisome.
 FT ACT SITE 65 BY SIMILARITY.
 FT ACT SITE 138 138 BY SIMILARITY.
 FT METAL 348 348 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT SITE 505 507 MICROBODY TARGETING SIGNAL (POTENTIAL).
 SQ SEQUENCE 507 AA; 57849 MW; 35366DDA49539CC3 CRC64;

Query Match 58.5%; Score 38; DB 1; Length 507;
 Best Local Similarity 70.0%; Pred. No. 68;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVGLSPGEQE 13

DB 445 VLGRTFGEQE 454

RESULT 13

C4AD DROME STANDARD; PRT; 516 AA.
 AC Q9VAT3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable cytochrome P450 4adi (EC 1.14.-.-) (CYP4AD1).
 GN CYP4AD1 OR CG2110.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George K.G., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celnik S.E.;
 RT "A Drosophila full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 CC -!- FUNCTION: May be involved in the metabolism of insect hormones and
 CC in the breakdown of synthetic insecticides (By similarity).
 CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
 CC oxidized flavoprotein + H(2)O.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
 CC (Potential).
 CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AB003837; AAF59092.1; --
 DR EMBL; AY061058; AAL28606.1; --
 DR HSSP; P14779; 13PZ.
 DR Flybase; FBgn0033292; Cyp4ad1.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
 KW Endoplasmic reticulum; Hypothetical protein.
 FT METAL 445 445 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 SQ SEQUENCE 516 AA; 58870 MW; 648EA22492AF58C7 CRC64;

Query Match 58.5%; Score 38; DB 1; Length 516;
 Best Local Similarity 80.0%; Pred. No. 69;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LPVAVGLSPG 10

DB 469 LPVAVGLSPG 478

RESULT 14

EFG HAEIN
 ID EFG HAEIN STANDARD; PRT; 699 AA.
 AC P43925;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-NOV-2003 (Rel. 41, Last annotation update)
 DE Elongation factor G (EF-G).
 GN FUSA OR FUS OR HI0579.

```

OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RG / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Ed.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC the nascent protein chain from the A-site to the P-site of the
CC ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
CC EF-G/EF-2 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U32739; AAC22237.1; -.
DR PIR; F64078; F64078.
DR HSSP; F13551; IELO.
DR TIGR; H10579; -.
DR HAMAP; MF_00054; -.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000640; EF_G.
DR InterPro; IPR009022; EFG_III_V.
DR InterPro; IPR005517; EFIV.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR005225; Small_GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00679; EF-G; 1.
DR Pfam; PF03764; EFIV; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR TIGRFAMs; TIGR00484; EF-G; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT INIT_MET 0 0 BY SIMILARITY.
FT NP_BIND 16 23 GTP (BY SIMILARITY).
FT NP_BIND 87 91 GTP (BY SIMILARITY).
FT NP_BIND 141 144 GTP (BY SIMILARITY).
SQ SEQUENCE 699 AA; 77132 MW; PBEADE39C0F62801 CRC64;

Query Match 58.5%; Score 38; DB 1; Length 699;
Best Local Similarity 46.2%; Pred. No. 93;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LPAVGLSPGEQE 13
Db 291 IPAIKGINPDETE 303

```

```

RESULT 15
ID EFG_PASMU STANDARD; PRT; 700 AA.
AC P57938;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Elongation factor G (EF-G).
GN FUSA OR PM1356.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RA "Complete genomic sequence of Pasteurella multocida Pm70.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC the nascent protein chain from the A-site to the P-site of the
CC ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the GTP-binding elongation factor family.
CC EF-G/EF-2 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE006173; AAK03440.1; -.
DR HSSP; P13551; IELO.
DR HAMAP; MF_00054; -.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000640; EF_G.
DR InterPro; IPR009022; EFG_III_V.
DR InterPro; IPR005517; EFG_IV.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR005225; Small_GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF03764; EFG_IV; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR TIGRFAMs; TIGR00484; EF-G; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT NP_BIND 16 23 GTP (BY SIMILARITY).
FT NP_BIND 87 91 GTP (BY SIMILARITY).
FT NP_BIND 141 144 GTP (BY SIMILARITY).
SQ SEQUENCE 700 AA; 77186 MW; 6CC161F7F9FA9C72 CRC64;

Query Match 58.5%; Score 38; DB 1; Length 700;
Best Local Similarity 46.2%; Pred. No. 93;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LPAVGLSPGEQE 13
Db 292 IPAIKGINPDETE 304

```

Search completed: May 7, 2004, 12:34:33
Job time: 4.68 secs

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	51	39.2	426	2	D72635	hypothetical prote
2	50.5	38.8	615	2	A3554	galactonate dehydr
3	50	38.5	277	2	F82130	formyltetrahydrofo
4	50	38.5	381	2	F75270	cytochrome P450 -
5	49.5	38.1	603	2	A52952	dihydroxy-acid deh
6	49.5	38.1	603	2	S38330	hypothetical prote
7	49	37.7	118	2	S38491	19 heavy chain - h
8	49	37.7	278	2	E84131	formyltetrahydrofo
9	48.5	37.3	559	2	S82503	inorganic phosphat
10	48	36.9	522	2	T44369	pyruvate, water di
11	48	36.9	567	2	AF0274	ribulokinase (EC 2
12	48	36.9	780	2	D75361	phosphoenolpyruvat
13	48	36.9	792	2	B82752	penicillin binding
14	47	36.2	280	2	F30845	formyltetrahydrofo
15	47	36.2	280	2	E85703	hypothetical prote
16	47	36.2	280	2	C36871	formyltetrahydrofo
17	47	36.2	280	2	AF0649	formyltetrahydrofo
18	47	36.2	293	2	E71819	formyltetrahydrofo
19	47	36.2	293	2	B64699	formyltetrahydrofo
20	47	36.2	661	2	G84511	formyltetrahydrofo
21	47	36.2	2109	1	I50421	hypothetical prote
22	46.5	35.8	178	2	G82977	aggreacan precursor
23	46.5	35.8	563	2	T36580	hypothetical prote
24	46.5	35.8	656	2	B81692	hypothetical prote
25	46	35.4	192	2	A82545	penicillin-binding
26	46	35.4	271	2	C92841	hypothetical prote
27	46	35.4	274	2	F81350	cysteine proteinas
28	46	35.4	287	2	F70871	formyltetrahydrofo
29	46	35.4	300	2	C69857	hypothetical prote

```

Query Match      38.8%; Score 50.5; DB 2; Length 615;
Best Local Similarity 45.8%; Pred. No. 26;
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 4 LPAVVGSLSP-----GQEQYHRRGV 22
Db 324 IPLLNVNLPAGEYLIGEDYHAGGV 347

RESULT 3
F82130
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82130
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: F82130
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-277 <HEI>
A:Cross-references: GB:AE004274; GB:AE003852; NID:g956533; PIDN:AAF95140.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1992
A:Map position: 1
C:Superfamily: phosphoribosylglycinamide formyltransferase; phosphoribosylglycinamide fo

Query Match      38.5%; Score 50; DB 2; Length 277;
Best Local Similarity 52.9%; Pred. No. 13;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 LPAVVGSLSPGQEQYHRRG 20
Db 190 LPAFIGAKPYQAYERG 206

RESULT 4
F75270
C:Species: Deinococcus radiodurans (strain R1)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: F75270
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75270
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-381 <WHI>
A:Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PID:g646029
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2473
A:Map position: 1

Query Match      38.5%; Score 50; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GLPAVVGSLSP 12
Db 51 GLPAVVGSLSP 60

```

```

RESULT 5
A82952
C:Species: Agrobacterium tumefaciens (strain C58, Dupont)
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: A82952
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: A82577; MUID:21608550; PMID:11743193
A:Accession: A82952
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-603 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA44035.1; PID:GL7741597; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: ilvD
A:Map position: linear chromosome
C:Superfamily: dihydroxy-acid dehydratase

Query Match      38.1%; Score 49.5; DB 2; Length 603;
Best Local Similarity 45.8%; Pred. No. 36;
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 4 LPAVVGSLSP-----GQEQYHRRGV 22
Db 312 VPLLNVNLPAGEYLIGEDYHAGGV 335

RESULT 6
H98330
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: H98330
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A:Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H98330
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-603 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK90170.1; PID:GL5160173; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_3190
A:Map position: linear chromosome
C:Superfamily: dihydroxy-acid dehydratase

Query Match      38.1%; Score 49.5; DB 2; Length 603;
Best Local Similarity 45.8%; Pred. No. 36;
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 4 LPAVVGSLSP-----GQEQYHRRGV 22
Db 312 VPLLNVNLPAGEYLIGEDYHAGGV 335

RESULT 7
S38491
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 23-Jul-1999
C:Accession: S38491
R:Markes, J.D.; Ouweland, W.H.; Bye, J.M.; Finnern, R.; Gorick, B.D.; Voak, D.; Thorpe,
submitted to the EMBL Data Library, June 1993
A:Description: Human antibody fragments specific for human blood group antigens from a

```

A;Reference number: S38488

A;Accession: S38491

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-118 <MAR>

A;Cross-references: EMBL:Z23032; NID:G414029; PIDN:CAA80567.1; PID:G414030

A;Superfamily: immunoglobulin V region; immunoglobulin homology

C;Keywords: heterotrimer; immunoglobulin

F;15-97/Domain: immunoglobulin homology <IMV>

Query Match 37.7%; Score 49; DB 2; Length 118;

Best Local Similarity 50.0%; Pred. No. 7.3;

Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 3 GLPAVVGSLSPGQYHRGV 22

Db 44 GLEWVGIGFGGDTYFGSV 53

RESULT 8

E64131

formyltetrahydrofolate deformylase (EC 3.5.1.10) - Haemophilus influenzae (strain Rd KW2)

C;Species: Haemophilus influenzae

C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 17-Mar-2000

C;Accession: E64131; FN0606

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.W.; Weidman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: E64131

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-278 <TYGR>

A;Cross-references: GB:U32833; GB:I42023; NID:G1574432; PIDN:AAAC23236.1; PID:G1574433; T

R;Maskell, D.

Gene 129, 155-156, 1993

A;Title: Cloning and sequencing of the Haemophilus influenzae aroA gene.

A;Reference number: JN0758; MUID:93328119; PMID:8335255

A;Accession: FN0606

A;Molecule type: DNA

A;Residues: 64-114, 'FNR', 118-137, 'PK', 141-204, 'E', 206-278 <MAS>

A;Cross-references: GB:L04686; NID:G148863; PIDN:AAA24942.1; PID:G148864

C;Genetics:

A;Gene: purN

C;Function:

A;Description: catalyzes hydrolysis of 10-formyltetrahydrofolate to tetrahydrofolate and

A;Pathway: one-carbon metabolism

A;Note: activated by methionine and inhibited by glycine

C;Superfamily: phosphoribosylglycinamide formyltransferase

C;Keywords: hydrolase

F;85-277/Domain: phosphoribosylglycinamide formyltransferase homology <PRGF>

F;223/Active site: Asp #status Predicted

Query Match 37.7%; Score 49; DB 2; Length 278;

Best Local Similarity 52.9%; Pred. No. 18;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 LPAVGLSPGQYHRG 20

Db 191 LPAFIGAKPYQAYKRG 207

RESULT 9

S62503

inorganic phosphate transporter - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 10-Dec-1999

C;Accession: T38287; S62503

R;Niblett, D.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, October 1995

A;Reference number: Z21783

A;Accession: T38287

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-559 <NI2>

A;Cross-references: EMBL:Z64354; NID:G1039338; PIDN:CAA91247.1; PID:G1039350; GSPDB:GNO

A;Experimental source: strain 972h; cosmid C23D3

C;Genetics:

A;Gene: SPDB:SPAC23D3.12

A;Map position: 1R

C;Superfamily: probable inorganic phosphate transport protein PHO84

Query Match 37.3%; Score 48.5; DB 2; Length 559;

Best Local Similarity 47.6%; Pred. No. 46;

Matches 10; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 8 VQLSPGQYH---RGVGV 25

Db 365 IGFSKGNEYHMLRGAINL 385

RESULT 10

T44369

pyruvate, water dikinase (EC 2.7.9.2) [imported] - Deinococcus radiodurans (fragment)

C;Species: Deinococcus radiodurans

C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 03-Jun-2002

C;Accession: T44369

R;Narumi, I.; Islam, S.; Cherdchu, K.; Kikuchi, M.; Watanabe, H.; Kitayama, S.; Yamamoto

submitted to the EMBL Data Library, August 1998

A;Description: I8301: the second insertion sequence element from Deinococcus radiodura

A;Reference number: Z22755

A;Accession: T44369

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-522 <NAR>

A;Cross-references: EMBL:AB016903; PIDN:BAAC32387.1

A;Experimental source: strain KD8301

C;Genetics:

A;Note: ppsA

C;Superfamily: Escherichia coli pyruvate, water dikinase; phosphotransferase system enzy

C;Keywords: transferase

Query Match 36.9%; Score 48; DB 2; Length 522;

Best Local Similarity 50.0%; Pred. No. 50;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 GLPAVVGSLSPGQYHRG 20

Db 160 GIPAVVGTCNATRELNG 177

RESULT 11

AF0274

ribulokinase (EC 2.7.1.16) [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001

C;Accession: AF0274

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, K.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AF0274

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-567 <KUR>

A;Cross-references: GB:AL590842; PIDN:CAC91058.1; PID:G15980250; GSPDB:GN00175

C;Genetics:

A;Gene: araB

C;Superfamily: ribulokinase

C;Keywords: phosphotransferase

Query Match 36.9%; Score 48; DB 2; Length 567;
Best Local Similarity 52.4%; Pred. No. 55;
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 3 GLPAVVGLSPGEQEVHRGGVG 23
|||||
Db 270 GLPTVILVSGAFDCHMGAVG 290
|||||

RESULT 12
D75361
phosphoenolpyruvate synthase - Deinococcus radiodurans (strain RL)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: D75361
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathhevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RL.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: D75361
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-780 <WHI>
A:Cross-references: GB:AE002014; GB:AE000513; NID:g6459494; PIDN:AAF11283.1; PID:g6459494
A:Experimental source: strain RL
C:Genetics:
A:Gene: DR1727
A:Map position: 1
C:Superfamily: Escherichia coli pyruvate, water dikinase; phosphotransferase system enzym

Query Match 36.9%; Score 48; DB 2; Length 780;
Best Local Similarity 50.0%; Pred. No. 77;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 GLPAVVGLSPGEQEVHRG 20
|:|:|:|:|:|:|
Db 418 GIPAVVGTGNATREHNG 435
|:|:|:|:|:|:|

RESULT 13
B82752
Penicillin binding protein 1B XF0884 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: B82752
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: B82752
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-792 <SIM>
A:Cross-references: GB:AE003927; GB:AE003849; NID:g9105783; PIDN:AAF83694.1; GSPDB:GN001
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
As-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.U.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 9 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-5

Perfect score: 130

Sequence: 1 MAGLPAVGLSPGEQYHGGVGL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	49	37.7	278	1	PURU_HAEIN
2	48.5	37.3	559	1	YABC_SCHPO
3	48	36.9	563	1	ARAB_YERPE
4	48	36.9	780	1	PEPSA_DEIRA
5	47	36.2	280	1	PURU_ECOL6
6	47	36.2	280	1	PURU_ECOLI
7	47	36.2	567	1	ARAB_VIBPA
8	47	36.2	813	1	CADM_MOUSE
9	47	36.2	2109	1	PGCA_CHICK
10	46	35.4	508	1	COBI_MYCTU
11	46	35.4	1172	1	CNA2_MOUSE
12	45	34.6	479	1	PTSB_VIBAL
13	45	34.6	961	1	ACON_MYCAV
14	44.5	34.2	618	1	ILVD_CHRVO
15	44.5	34.2	2054	1	M18A_HUMAN
16	44	33.8	151	1	SODC_XIPGL
17	44	33.8	433	1	ENOB_RABIT
18	44	33.8	565	1	ARAB_ECO57
19	44	33.8	565	1	ARAB_ECOL6
20	44	33.8	565	1	ARAB_ECOLI
21	44	33.8	568	1	ARAB_SALTI
22	44	33.8	614	1	CPRI_DROME
23	44	33.8	617	1	ILVD_STPCO
24	44	33.8	658	1	VG18_BP74
25	44	33.8	677	1	SG1_HUMAN
26	44	33.8	777	1	L100_ADE41
27	44	33.8	813	1	CADM_RAT
28	44	33.8	1065	1	RPOB_MARPO
29	44	33.8	3664	1	MINT_HUMAN
30	44	33.8	4351	1	PAT2_RAT
31	43.5	33.5	173	1	RUVG_XYLFA
32	43.5	33.5	173	1	RUVG_XYLFT
33	43.5	33.5	968	1	CTDI_HUMAN

34	43	33.1	211	1	COBL_METJA
35	43	33.1	325	1	CA19_RAT
36	43	33.1	326	1	MER_METTI
37	43	33.1	364	1	COAL_SV40
38	43	33.1	382	1	FETE_HUMAN
39	43	33.1	390	1	COBL_MYCTU
40	43	33.1	481	1	GLGA_RHILO
41	43	33.1	505	1	Y76J_CAEEL
42	43	33.1	516	1	C4AD_DROME
43	43	33.1	544	1	CH60_PROAC
44	43	33.1	561	1	LCFA_SALTY
45	43	33.1	628	1	HNFA_MOUSE

ALIGNMENTS

RESULT 1
PURU_HAEIN
ID PURU_HAEIN STANDARD; PRT; 278 AA.
AC Q03432;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4) hydrolase).
DE hydrolase).
GN PURU OR H11588.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RG / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kieravage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Ghem C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
RL "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. ";
RL Science 269:496-512(1995).
RN [2]
RP SEQUENCE OF 64-278 FROM N.A.
RC STRAIN=RM 7004 / Serotype B;
RX MEDLINE=93328119; PubMed=8335255;
RA Maskell D.J.;
RL "Cloning and sequencing of the Haemophilus influenzae aroA gene. ";
RL Gene 129:153-156(1993).
CC -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate + tetrahydrofolate.
CC -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine (By similarity).
CC -!- PATHWAY: De novo purine biosynthesis.
CC -!- SUBUNIT: Homohexamer (By similarity).
CC -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

O58917	methanococc
P20850	rattus norv
Q9uxp0	methanolobu
P03087	simian viru
Q9ugm5	homo sapien
Q10671	mycobacteri
Q985p2	rhizobium l
P90938	caenorhabdi
Q9v4t3	drosophila
Q9k2u4	propionibac
Q8xg98	salmonella
P22361	mus musculu


```

or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U22833; AAC23236.1; -.
CC EMBL; L04686; AAA24942.1; -.
CC PIR; E64131; E64131.
CC HSP; P08179; IGRC.
CC TIGR; H1588; -.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR002376; formyl_transf.
CC InterPro; IPR004810; PurU.
CC Pfam; PF01842; ACT; 1.
CC Pfam; PF00551; formyl_transf; 1.
CC PRINTS; PR01575; FFH4YDLASE.
CC TIGRfam; TIGR00655; PurU; 1.
CC Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW Complete proteome.
FT ACT_SITE 223 BY SIMILARITY.
FT CONFLICT 115 117 VIG -> RNR (IN REF. 2).
FT CONFLICT 138 140 HEN -> PK (IN REF. 2).
FT CONFLICT 205 205 K -> E (IN REF. 2).
SQ SEQUENCE 278 AA; 32173 MW; 7F375AB3C422EC4B CRC64;

Query Match 37.7%; Score 49; DB 1; Length 278;
Best Local Similarity 52.9%; Pred. No. 8.6;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 LPVVGLSPGEQYHRG 20
DB 191 LPAFIGARPYQAYKRG 207

RESULT 2
ID VAEC SCHPO STANDARD; PRT; 559 AA.
AC Q09852;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative inorganic phosphate transporter C3D3.12.
GN SPAC2D3.12.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Murphy L., Niblett D., Odell C.,
RA Rutherford K., Rutter S., Saunders D., Quail M.A., Rabinowitch E.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicakert G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

```

RT The genome sequence of Schizosaccharomyces pombe. ";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: High-affinity transporter for external inorganic
CC phosphate (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. STRONG, TO
CC YEAST PHO84.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z64354; CAA91247.1; -.
CC PIR; T38287; S62503.
CC GeneDB_Spombe; SPAC2D3.12; -.
CC InterPro; IPR007114; MFS.
CC InterPro; IPR004738; Phos_permease.
CC InterPro; IPR005828; Sub_transporter.
CC InterPro; IPR005829; Sug_transporter.
CC Pfam; PF00083; sugar_tr; 1.
CC TIGRfam; TIGR00887; 2A0109; 1.
CC PROSITE; PS00850; MFS; 1.
CC PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
CC PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
CC KW Hypothetical protein; Phosphate transport; Transport; Transmembrane.
FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 46 66 1 (POTENTIAL).
FT DOMAIN 67 94 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 95 115 2 (POTENTIAL).
FT DOMAIN 116 118 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 119 139 3 (POTENTIAL).
FT DOMAIN 140 144 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 145 165 4 (POTENTIAL).
FT DOMAIN 166 188 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 189 209 5 (POTENTIAL).
FT DOMAIN 210 224 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 225 245 6 (POTENTIAL).
FT DOMAIN 246 335 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 336 356 7 (POTENTIAL).
FT DOMAIN 357 381 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 382 402 8 (POTENTIAL).
FT DOMAIN 403 409 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 410 430 9 (POTENTIAL).
FT DOMAIN 431 438 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 439 459 10 (POTENTIAL).
FT DOMAIN 460 474 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 475 495 11 (POTENTIAL).
FT DOMAIN 496 497 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 498 518 12 (POTENTIAL).
FT DOMAIN 519 559 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 559 AA; 62316 MW; 2E8958F86C2092E2 CRC64;

Query Match 37.3%; Score 48.5; DB 1; Length 559;
Best Local Similarity 47.6%; Pred. No. 21;
Matches 10; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 8 VGLSPGEQYH---RGVGVVL 25
DB 365 IGFSGKNEYHTLMRGAINL 385

RESULT 3
ID ARAB_YERPE STANDARD; PRT; 563 AA.
AC P58543;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE L-ribulokinase (EC 2.7.1.16).

```

```

GN ARAB OR YPO2254 OR Y2095.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=CO-92 / Biovar Orientalis;
RC MEDLINE=21470413; PubMed=11586360;
RX Parhill J., Wren B.M., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Partridge M.B., Sebahia M., James K.D., Churcher C., Mungall K.I.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth I., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]_SEQUENCE FROM N.A.
RP STRAIN=KIM5 / Biovar Mediaevalis;
RC MEDLINE=22137863; PubMed=12142430;
RX Deng W., Hurland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
RN [3]_CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
phosphate.
CC -!- PATHWAY: L-arabinose catabolism; second step.
CC -!- SIMILARITY: Belongs to the ribulokinase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ414151; CAC91058.1; ALT INIT.
DR EMBL; AE013812; AAM85658.1; ALT_INIT.
DR HAMAP; MF 00520; -.
DR InterPro; IPR000577; FGGY_kin.
DR InterPro; IPR005929; L-ribulokin.
DR Pfam; PF00370; FGGY; 1.
DR Pfam; PF02782; FGGY_C; 1.
DR TIGRFAMs; TIGR01234; L-ribulokinase; 1.
DR Transferase; Kinase; Arabinose catabolism; Complete proteome.
SQ SEQUENCE 563 AA; 61833 MW; 88EF1E030A5414E CRC64;

Query Match 36.9%; Score 48; DB 1; Length 563;
Best Local Similarity 52.4%; Pred. No. 25;
Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 3 GLPAAVVGSLPGGEYHRRGGV 23
Db 266 GLPTTVILSGGAPDCHMGAVG 286

RESULT 4
ID -PPSA_DEIRA STANDARD; PRT; 780 AA.
AC O83026;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate, water dikinase)
DE (PEP synthase).
GN -PPSA OR DR1727.
OS Deinococcus radiodurans.

```

```

OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=RL / ATCC 13939 / DSM 20539 / NCIB 9279;
RC MEDLINE=20036896; PubMed=10567266;
RX White O., Eison J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Ueberback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RL Science 286:1571-1577(1999).
RN [2]_SEQUENCE OF 259-780 FROM N.A.
RP STRAIN=I., Islam S., Cherdchu K., Kikuchi M., Watanabe H.,
RA Kitayama S., Yamamoto K.;
RT "I8301: the second insertion sequence element from Deinococcus
RT radiodurans.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
phosphoenolpyruvate + phosphate.
CC -!- PATHWAY: ESSENTIAL STEP IN GLUCONEOGENESIS WHEN PYRUVATE AND
LACTATE ARE USED AS A CARBON SOURCE.
CC -!- SIMILARITY: Belongs to the PEP-utilizing enzyme family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002014; AAF11283.1; -.
DR EMBL; AB016803; BAA32387.1; -.
DR PIR; D75361; D75361.
DR PIR; T44369; T44369.
DR TIGR; DR1727; -.
DR InterPro; IPR008279; PEP_mobile.
DR InterPro; IPR006318; PEP_P_trans.
DR InterPro; IPR006319; PEP_synth.
DR InterPro; IPR000121; PEP_utilizers.
DR InterPro; IPR002192; PPD_K_N_term.
DR Pfam; PF00391; PEP-utilizers; 1.
DR Pfam; PF02896; PEP-utilizers_C; 1.
DR Pfam; PF01326; PPD_K_N_term; 1.
DR PRINTS; PR01736; PHPTRNPRASE.
DR ProDom; PD000940; PEP_utilizers; 1.
DR TIGRFAMs; TIGR01418; PEP_synth; 1.
DR PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR PROSITE; PS00742; PEP_ENZYMES_2; 1.
KW Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.
FT ACT_SITE 409 409 TELE-PHOSPHOHISTIDINE INTERMEDIATE
FT MOD_RES 409 409 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 780 AA; 84895 MW; AD555076324ADA47 CRC64;

Query Match 36.9%; Score 48; DB 1; Length 780;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 GLPAAVVGSLPGGEYHRRG 20
Db 418 GIPAAVVGATRELHNG 435

RESULT 5
PURI_ECOL6

```

ID PURU ECOL6 STANDARD; PRT; 280 AA.
 AC P38480;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
 DE hydrolase).
 GN PURU OR C1696 OR SF1232 OR S1318.
 OS Escherichia coli O6, and
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992, 623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINE=93023838; PubMed=1406252;
 RA Hromockyj A.E., Tucker S.C., Maurelli A.T.;
 RT "Temperature regulation of Shigella virulence: identification of the
 RT repressor gene virR, an analogue of hns, and partial complementation
 RT by tyrosyl transfer RNA (tRNA^{Tyr}).";
 RL Mol. Microbiol. 6:2113-2124 (1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 RT through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441 (2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.;
 RT "Complete genome sequence and comparative genomics of Shigella
 RT flexneri serotype 2a strain 2457T.";
 RL Infect. Immun. 71:2775-2786 (2003).
 RN [5]
 RP IDENTIFICATION
 RX MEDLINE=94042872; PubMed=8226647;
 RA Nagy P.L., McCorkle G., Zalkin H.;
 RT "puru, a source of formate for purT-dependent phosphoribosyl-N-
 RT formylglycinamide synthesis.";
 RL J. Bacteriol. 175:7066-7073 (1993).
 CC -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
 CC 5'-PHOSPHORIBOSYL-N-FORMYLGLYCNAMIDE (FGAR) DURING AEROBIC
 CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.
 CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H₂O = formate +
 CC tetrahydrofolate.
 CC -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine
 CC (By similarity).
 CC -!- PATHWAY: De novo purine biosynthesis.
 CC -!- SUBUNIT: Homohexamer (By similarity).

CC -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB016760; AAN80163.1; -;
 DR EMBL; X66849; -; NOT ANNOTATED_CDS.
 DR EMBL; AB015150; AAN42845.1; -;
 DR EMBL; AB016982; RAP16730.1; -;
 DR HSSP; P08179; IGRG.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR002376; formyl_transf.
 DR InterPro; IPR004810; PurU.
 DR Pfam; PF01842; ACT; 1.
 DR Pfam; PF00551; formyl_transf; 1.
 DR PRINTS; PR01575; FFH4HYDRLASE.
 DR TIGRfams; TIGR00655; PurU; 1.
 KW Purine biosynthesis; Hydrolase; One-carbon metabolism;
 KW Complete proteome.
 FT ACCT SITE 225 225 BY SIMILARITY.
 FT CONFLICT 44 44 R -> L (IN REF. 2).
 SQ SEQUENCE 280 AA; 31920 MW; 55BC16B62727A419 CRC64;
 Query Match 36.2%; Score 47; DB 1; Length 280;
 Best Local Similarity 52.9%; Pred. No. 17;
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 OY 4 LPVVGLSPGEYHVG 20
 DB 193 LPAFGARPYHQAVERG 209
 RESULT 6
 ID PURU ECOLI STANDARD; PRT; 280 AA.
 AC P37051;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
 DE hydrolase).
 GN PURU OR TGS OR B1232.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.
 RC STRAIN=K12;
 RX MEDLINE=94042872; PubMed=8226647;
 RA Nagy P.L., McCorkle G., Zalkin H.;
 RT "puru, a source of formate for purT-dependent phosphoribosyl-N-
 RT formylglycinamide synthesis.";
 RL J. Bacteriol. 175:7066-7073 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94110230; PubMed=8282700;
 RA Boesl M., Kersten H.;
 RT "Organization and functions of genes in the upstream region of tyrT
 RT of Escherichia coli: phenotypes of mutants with partial deletion of a
 RT new gene (tgs).";
 RL J. Bacteriol. 176:221-231 (1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474 (1997).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=9061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-Kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155 (1996).
 [5]
 RP CHARACTERIZATION.
 RX MEDLINE=95173107; PubMed=7868604;
 RA Nagy P.L., Marolewski A., Benkovic S.J., Zalkin H.;
 RT "Formyltetrahydrofolate hydrolase, a regulatory enzyme that functions
 to balance pools of tetrahydrofolate and one-carbon tetrahydrofolate
 adducts in *Escherichia coli*.";
 RL J. Bacteriol. 177:1292-1298 (1995).
 CC -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
 CC THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
 CC 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
 CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.
 CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
 CC tetrahydrofolate.
 CC -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine.
 CC -!- PATHWAY: De novo purine biosynthesis.
 CC -!- SUBUNIT: Homohexamer.
 CC -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L20251; AAC36846.1; -;
 CC EMBL; M64675; AAA16860.1; ALT INIT.
 CC EMBL; AE000221; AAC74314.1; -;
 CC EMBL; D90758; BAA36100.1; -;
 CC EMBL; D90759; BAA36112.1; -;
 CC EMBL; D90852; BAA16026.1; -;
 CC PIR; C36871; C36871.
 CC EcGene; EG11819; purU.
 CC InterPro; IPR002912; ACT.
 CC InterPro; IPR002376; formyl_transf.
 CC InterPro; IPR004810; PurU.
 CC Pfam; PF01842; ACT; 1.
 CC Pfam; PF00551; formyl_transf; 1.
 CC PRINTS; PR01575; FFH4HYDRLASE.
 CC TIGRfams; TIGR00655; PurU; 1.
 CC Purine biosynthesis; Hydrolase; One-carbon metabolism;
 CC Complete proteome.
 CC ACT SITE 225 225 BY SIMILARITY.
 CC SEQUENCE 280 AA; 31934 MW; 5667406D2727A2C2 CRC64;
 Query Match 36.2%; Score 47; DB 1; Length 280;
 Best Local Similarity 52.9%; Pred.No. 17;
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 QY 4 LPAAVGLSPGQEQYHRG 20
 ||| : : : : :
 DB 193 LPAAFGARPHQAYERG 209

RESULT 7
 ARAB_VIBPA STANDARD; PRT; 567 AA.
 ID ARAB_VIBPA
 AC Q87FK5;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE L-ribulokinase (EC 2.7.1.16).
 GN ARAB OR VPA1674.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RMID 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 distinct from that of *V. cholerae*.";
 RL Lancet 361:743-749 (2003).
 CC -!- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
 CC phosphate.
 CC -!- PATHWAY: L-arabinose catabolism; second step.
 CC -!- SIMILARITY: Belongs to the ribulokinase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AP005089; BAC63017.1; -;
 CC HAMAP; MF_00520; -; i.
 CC InterPro; IPR000577; FGGY_kin.
 CC Pfam; PF00370; FGGY; 1.
 CC Pfam; PF02782; FGGY_C; 1.
 CC Transferase; Kinase; Arabinose catabolism; Complete proteome.
 CC SEQUENCE 567 AA; 61788 MW; B303B547EBEA9B8B CRC64;
 Query Match 36.2%; Score 47; DB 1; Length 567;
 Best Local Similarity 42.9%; Pred.No. 35;
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
 QY 3 GLPAVVLSPGQEQYHRGGVG 23
 ||| : : : : :
 DB 267 GLPEGTAIGAEGFDCHMGAVG 287
 RESULT 8
 CADM_MOUSE STANDARD; PRT; 813 AA.
 ID CADM_MOUSE
 AC Q8WT5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cadherin-22 precursor (PB-cadherin).
 GN CDH22.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
 RC STRAIN=ICR; TISSUE=Brain;
 RX MEDLINE=99326347; PubMed=10398531;
 RA Kitajima K., Koshimizu U., Nakamura T.;
 RT "Expression of a novel type of classic cadherin, PB-cadherin in

RT developing brain and limb buds. ";
 RL Dev. Dyn. 215:206-214(1999).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types. P3-cadherins may have a role
 CC in the morphological organization of pituitary gland and brain
 CC tissues.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Predominantly expressed in brain. Abundant in
 CC olfactory bulb, cerebellum, and cerebellum, less in pons, medulla,
 CC and spinal cord. Low expression in heart. No expression in lung,
 CC liver, spleen, kidney, testis, stomach, intestine, colon, and
 CC placenta.
 CC -!- DEVELOPMENTAL STAGE: Expressed at 9.5 dpc onwards. At 10.5 dpc, in
 CC brain (telencephalic vesicles and isthmus), spinal cord and limb
 CC buds (in the zone of polarizing activity). At 14.5 dpc, in
 CC olfactory bulb and cerebellum.
 CC -!- INDUCTION: Down-regulated by thyroid hormone.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB019618; BAA34426.1; --
 CC MGD; MGI:1341843; Cdh22.
 CC HSP; P15116; INCU.
 CC InterPro: IPR002126; Cadherin.
 CC InterPro: IPR000233; Cadherin_C term.
 CC Pfam; PF00028; cadherin; 5.
 CC Pfam; PF01049; Cadherin_C term; 1.
 CC PRINTS; PR00205; CADHERIN.
 CC SMART; SM00112; CA; 5.
 CC PROSITE; PS00232; CADHERIN_1; 2.
 CC PROSITE; PS00268; CADHERIN_2; 5.
 CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal.
 KW SIGNAL.
 FT CHAIN 1 33 POTENTIAL.
 FT CHAIN 34 813 CADHERIN-22.
 FT DOMAIN 33 621 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 622 642 POTENTIAL.
 FT DOMAIN 643 813 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 61 165 CADHERIN 1.
 FT DOMAIN 166 274 CADHERIN 2.
 FT DOMAIN 275 391 CADHERIN 3.
 FT DOMAIN 392 495 CADHERIN 4.
 FT DOMAIN 496 613 CADHERIN 5.
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 609 609 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 813 AA; 88021 MW; 5510F9848D976567 CRC64;
 Query Match 36.2%; Score 47; DB 1; Length 813;
 Best Local Similarity 50.0%; Pred. NO. 51;
 Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
 QY 2 AGLPAVGLSPGEQYHRGGV 23
 Db 34 ASTPAPSSLSPGAQDNQLGAG 55
 RESULT 9
 PGCA CHICK
 ID PGCA CHICK STANDARD; PRT; 2109 AA.
 AC P07898; Q90820; Q90991; Q91047;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
 DE protein) (CSPCP).
 GN AGC1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=White leghorn; TISSUE=Embryo;
 RX MEDLINE=94043149; PubMed=8226878;
 RA Li H., Schwartz N.B., Vertel B.M.;
 RT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core
 RT protein and identification of a stop codon in the aggrecan gene
 RT associated with the chondrodystrophy, nanomelia.";
 RL J. Biol. Chem. 268:23504-23511(1993).
 RN [2]
 RP SEQUENCE OF 1042-1559 FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=90307744; PubMed=1694853;
 RA Krueger R.C. Jr., Fields T.A., Mensch J.R. Jr., Schwartz N.B.;
 RT "Chick cartilage chondroitin sulfate proteoglycan core protein. II.
 RT Nucleotide sequence of cDNA clone and localization of the S103L
 RT epitope.";
 RL J. Biol. Chem. 265:12088-12097(1990).
 RN [3]
 RP SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
 RC TISSUE=Cartilage;
 RX MEDLINE=93111968; PubMed=1339285;
 RA Chandrasekaran L., Tanzer M.L.;
 RT "Molecular cloning of chicken aggrecan. Structural analyses.";
 RL Biochem. J. 288:903-910(1992).
 RN [4]
 RP ERRATUM.
 RX MEDLINE=94107258; PubMed=8280087;
 RA Chandrasekaran L., Tanzer M.L.;
 RL Biochem. J. 296:885-887(1993).
 RN [5]
 RP SEQUENCE OF 1492-1610 FROM N.A.
 RC STRAIN=White leghorn; TISSUE=Chondrocytes;
 RX MEDLINE=95129519; PubMed=7827752;
 RA Primorac D., Stover M.L., Clark S.H., Rowe D.W.;
 RT "Molecular basis of nanomelia, a heritable chondrodystrophy of
 RT chicken.";
 RL Matrix Biol. 14:297-305(1994).
 RN [6]
 RP SEQUENCE OF 1894-2109 FROM N.A.
 RX MEDLINE=89008506; PubMed=3170613;
 RA Tanaka T., Har-El R., Tanzer M.L.;
 RT "Partial structure of the gene for chicken cartilage proteoglycan
 RT core protein.";
 RL J. Biol. Chem. 263:15831-15835(1988).
 RN [7]
 RP SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.
 RX MEDLINE=86259736; PubMed=3460082;
 RA Sai S., Tanaka T., Koshier R.A., Tanzer M.L.;
 RT "Cloning and sequence analysis of a partial cDNA for chicken
 RT cartilage proteoglycan core protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).
 CC -!- FUNCTION: This proteoglycan is a major component of extracellular
 CC matrix of cartilaginous tissues. A major function of this protein
 CC is to resist compression in cartilage. It binds avidly to
 CC hyaluronic acid via an amino-terminal globular region. May play a
 CC regulatory role in the matrix assembly of the cartilage.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
 CC similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P07898-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P07898-2; Sequence=VSP_003073;

CC CC -1- DOMAIN: Two globular domains, G1 and G2, comprise the amino terminus of the proteoglycan, while another globular region, G3, makes up the COOH terminus. G1 contains link domains and thus consists of three disulfide-bonded loop structures designated as the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS) and the chondroitin sulfate (CS) attachment domains lie between G2 and G3.

CC CC -1- PTM: Contains mostly chondroitin sulfate, but also keratan sulfate chains, N-linked and O-linked oligosaccharides.

CC CC -1- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF NANOMELIA, A LETHAL CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT (CHONDRODYSPLASIA) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS. AGGREGAN IS TRUNCATED AT ITS C-TERMINAL IN THE CS-2 BINDING DOMAIN AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.

CC CC -1- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.

CC CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

CC CC -1- SIMILARITY: Contains 4 link domains.

CC CC -1- SIMILARITY: Contains 1 EGF-like domain.

CC CC -1- SIMILARITY: Contains 1 C-type lectin family domain.

CC CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.

CC CC -----

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC CC -----

DR EMBL; L21913; AAB19128.1; -;
DR EMBL; M38187; AAA48731.1; -;
DR EMBL; M88101; -; NOT ANNOTATED_CDS.
DR EMBL; S74657; AAC60751.1; -;
DR EMBL; S74656; AAC60751.1; JOINED.
DR EMBL; J04028; AAA48719.1; -;
DR EMBL; M13993; AAA48720.1; -;
DR PIR; I50421; I50421.
DR HSP; P08709; 1BF9.
DR InterPro; IPR002353; AntifreezeII.
DR InterPro; IPR000152; Asx hydroxyl_5.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR001304; Linkin_C.
DR InterPro; IPR000538; Linkin_C.
DR InterPro; IPR003324; SGXSG.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF02339; SGXSG; 56.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 4.
DR PRINTS; PR00356; ANTIFREEZEII.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 4.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; Ig; 1.
DR SMART; SM00445; LINK; 4.
DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00441; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS00026; EGF_3; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG LINK; 1.
DR PROSITE; PS01241; LINK; 4.
KW Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; EGF-like domain;
KW Alternative splicing; Repeat; Immunoglobulin domain.

FT SIGNAL	1	16	POTENTIAL.
FT CHAIN	17	2109	AGGREGAN CORE PROTEIN.
FT DOMAIN	34	143	IG-LIKE V-TYPE.
FT DOMAIN	166	243	LINK 1.
FT DOMAIN	264	346	LINK 2.
FT DOMAIN	537	614	LINK 3.
FT DOMAIN	635	716	LINK 4.
FT DOMAIN	1363	1742	19 X 20 AA TANDEM-REPEAT.
FT DOMAIN	1855	1892	EGF-LIKE.
FT DOMAIN	1901	2019	C-TYPE LECTIN.
FT DOMAIN	2023	2081	SUSHI.
FT DOMAIN	48	137	G1-A.
FT DOMAIN	148	243	G1-B.
FT DOMAIN	249	346	G1-B'.
FT DOMAIN	519	613	G2-B.
FT DOMAIN	620	715	G2-B'.
FT DOMAIN	718	803	KS.
FT DOMAIN	805	1264	CS-1.
FT DOMAIN	1265	1742	CS-2.
FT DOMAIN	1893	2109	G3.
FT DISULFID	51	129	BY SIMILARITY.
FT DISULFID	171	242	BY SIMILARITY.
FT DISULFID	195	216	BY SIMILARITY.
FT DISULFID	269	345	BY SIMILARITY.
FT DISULFID	293	344	BY SIMILARITY.
FT DISULFID	542	613	BY SIMILARITY.
FT DISULFID	566	587	BY SIMILARITY.
FT DISULFID	640	715	BY SIMILARITY.
FT DISULFID	664	685	BY SIMILARITY.
FT DISULFID	1859	1870	BY SIMILARITY.
FT DISULFID	1864	1879	BY SIMILARITY.
FT DISULFID	1881	1890	BY SIMILARITY.
FT DISULFID	1897	1908	BY SIMILARITY.
FT DISULFID	1925	2017	BY SIMILARITY.
FT DISULFID	1993	2009	BY SIMILARITY.
FT DISULFID	2024	2067	BY SIMILARITY.
FT DISULFID	2053	2080	BY SIMILARITY.
FT CARBOHYD	76	76	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	122	122	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	330	330	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	388	388	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	439	439	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	644	644	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	700	700	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	765	765	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	801	801	N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC	1856	1892	Missing (in isoform 2).
FT CONFLICT	362	362	/FTid=VSP 003073; E -> D (IN REF. 3).

Query Match 36.2%; Score 47; DB 1; Length 2109;
Best Local Similarity 47.6%; Pred. No. 1.4e+02;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 4 LPAYVGLSPGEQYHRGVGV 24

Db 957 LVEVVTAPGRQERKGSIGV 977

RESULT 10

COBI_MYCTU

ID COBI_MYCTU

AC Q10677; STANDARD; PRT; 508 AA.

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cobalamin biosynthesis protein COBI [includes: Precorrin-2 C20-

DE methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine-precorrin-2

DE methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-).]

GN COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.

OS Mycobacterium tuberculosis, and

OS Bacteria; Actinobacteria; Actinomycetales;

Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
 methyltransferase; Multifunctional enzyme; Complete proteome.
 1 243 PRECORRIN-2 C20-METHYLTRANSFERASE.
 DOWMAIN


```

DR InterPro; IPR003107; HAT.
DR SMART; SM00386; HAT; 5.
KW Repeat; Coiled coil.
FT DOMAIN 78 330 COILED COIL (POTENTIAL).
FT REPEAT 314 346 HAT 1.
FT REPEAT 404 436 HAT 2.
FT REPEAT 766 800 HAT 3.
FT REPEAT 986 1018 HAT 4.
FT REPEAT 1075 1109 HAT 5.
FT DOMAIN 86 93 POLY-LYS.
FT DOMAIN 146 154 POLY-ALA.
FT CONFLICT 707 707 R -> H (IN REF. 1; AAH25577).
FT CONFLICT 741 742 LS -> VC (IN REF. 1; AAH25577).
SQ SEQUENCE 1172 AA; 133424 MW; AB37A8BCE7A528F CRC64;

Query Match 35.4%; Score 46; DB 1; Length 1172;
Best Local Similarity 42.9%; Pred. No. 1e+02;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 5 PAVVGLSPGGEQVHRGVGV 25
DQ 351 PGVYALGEGEGEKHRSKL 371

RESULT 12
ID PTSB VIBAL STANDARD; PRT; 479 AA.
AC P22825;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE PTS system, sucrose-specific IIBC component (EIBC-SCR) (Sucrose-
DE permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=663;
RN [1]
RP MEDLINE=91071601; PubMed=2174811;
RX Blatch G.L., Scholle R.R., Woods D.R.;
RA "Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose
RT uptake-encoding region.";
RL Gene 95:17-23 (1990).
RN [2]
RP SEQUENCE OF 1-12 FROM N.A.
RX MEDLINE=91285433; PubMed=2060795;
RA Blatch G.L., Woods D.R.;
RT "Nucleotide sequence and analysis of the Vibrio alginolyticus scr
RL repressor-encoding gene (scrR)".
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91071601; PubMed=2174811;
RX Blatch G.L., Scholle R.R., Woods D.R.;
RA "Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose
RT uptake-encoding region.";
RL Gene 95:17-23 (1990).

-!- FUNCTION: This is a component of the phosphoenolpyruvate-dependent
sugar phosphotransferase system (PTS), a major carbohydrate active
-transport system. The IICD domains contain the sugar binding site
and the transmembrane channel; the IIA domain contains the primary
phosphorylation site (the donor is phospho-HPr); IIB transfers its
phosphoryl group to the IIB domain which finally transfers it to
the sugar.
-!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
histidine + sugar phosphate.
-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
-!- SIMILARITY: Contains 1 PTS EIIA domain.
-!- SIMILARITY: Contains 1 PTS EIIC domain.
-!- SIMILARITY: Contains 1 PTS EIIC domain.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AF002133; AAC46192.1;
InterPro; IPR006249; Aconitase_1.
InterPro; IPR000573; Aconitase_C.
InterPro; IPR001030; Aconitase_N.
Pfam; PF00330; aconitase; 1.
Pfam; PF00694; Aconitase_C; 1.

```

```

CC EMBL; M76768; AAA27555.1;
DR EMBL; M35009; AAA27557.2;
DR PIR; JQ0781; JQ0781.
DR HSSP; P05053; IIBA.
DR InterPro; IPR001996; Ptrans_EIIB.
DR InterPro; IPR003352; Ptrans_EIIC.
DR InterPro; IPR004719; PTSIIC_glc.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR TIGRFAMs; TIGR00826; EIIA_glc; 1.
DR TIGRFAMs; TIGR00852; pts-glc; 1.
DR PROSITE; PS01035; PTS_EIIB_CVS; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Transmembrane; Inner membrane; Phosphorylation.
FT DOMAIN 1 ? 479 EIIB.
FT MOD RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
FT MOD RES 324 324 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 479 AA; 49890 MW; FB906B5170E3EB7 CRC64;

Query Match 34.6%; Score 45; DB 1; Length 479;
Best Local Similarity 33.3%; Pred. No. 58;
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 2 AGLPAVGLSPGGEQVHRGV 22
DQ 431 AGLPGIISITPKIGYIAGM 451

RESULT 13
ACON MYCAV
ID ACON MYCAV STANDARD; PRT; 961 AA.
AC O08451;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aconitate hydratase (EC 4.2.1.3) (Citrate hydro-lyase) (Aconitase).
GN ACN.
OS Mycobacterium avium.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1764;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GJ10;
RX MEDLINE=98195739; PubMed=9534249;
RA Labo M., Guesbert L., de Rossi E., Speziale P., Riccardi G.;
RT "Determination of a 15437 bp nucleotide sequence around the inhA gene
of Mycobacterium avium and similarity analysis of the products of
putative ORFs.";
RN [2]
RP Microbiology 144:807-814 (1998).
CC -!- CATALYTIC ACTIVITY: Citrate = cis-aconitate + H(2)O.
CC -!- COFACTOR: Binds 1 4Fe-4S cluster per subunit (By similarity).
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: Belongs to the aconitase/IPM isomerase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AF002133; AAC46192.1;
InterPro; IPR006249; Aconitase_1.
InterPro; IPR000573; Aconitase_C.
InterPro; IPR001030; Aconitase_N.
Pfam; PF00330; aconitase; 1.
Pfam; PF00694; Aconitase_C; 1.

```


DR PRINTS; PR00415; ACONITASE.
 DR PRODOM; PD000511; Aconitase N; 1.
 DR TIGRFAMs; TIGR01341; aconitase 1; 1.
 DR PROSITE; PS00450; ACONITASE_1; 1.
 DR PROSITE; PS01244; ACONITASE_2; 1.
 KW Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S.
 FT METAL 499
 FT METAL 565
 FT METAL 568
 SQ SEQUENCE 961 AA; 104025 MW; 689566D95E31F596 CRC64;
 Query Match 34.6%; Score 45; DB 1; Length 961;
 Best Local Similarity 48.0%; Pred. No. 1.2e+02;
 Matches 12; Conservative 5; Mismatches 6; Indels 2; Gaps 2;
 QY 2 AGLPAVVGL-SPGEQYHRCGVGL 25
 Db 927 AETDAVVRITPGEADYIRNG-GIL 950
 RESULT 14
 ILVD CHRVQ
 ID ILVD CHRVQ STANDARD; PRT; 618 AA.
 AC Q7NYJ7;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Dihydroxy-acid dehydratase [EC 4.2.1.9] (DAD).
 GN ILVD OR CV1277.
 OS Chromobacterium violaceum.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Chromobacterium.
 ON NCBI_TaxID=536;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12472 / DSM 30191;
 RX MEDLINE=22882880; PubMed=14500782;
 RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
 RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
 RA Alves-Gomes J.A., Andrade E.M., Azarpe J., de Araujo M.F.F.,
 RA Astolfi-Filho S., Azevedo V., Baptista A.J., Batais L.A.M.,
 RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
 RA Boudignon J., Brigidio M.M., Brito C.A., Broochi M., Burity H.A.,
 RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carrao M.,
 RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
 RA Creczynski-pasa T.B., Cunha-Junior N.C., Fagundes N., Faicao C.L.,
 RA Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
 RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furian L.R.,
 RA Garzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
 RA Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,
 RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
 RA Madeira H.M.F., Manfio G.P., Maranhao A.O., Martins W.S.,
 RA di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,
 RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
 RA Paixao R.F.C., Parente J.A., Piedra F.O., Pena S.D.J., Pereira J.O.,
 RA Pereira M., Pinto L.S.R.C., Pintos L.S., Porto J.I.R., Potrich D.P.,
 RA Rantolho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
 RA Santos E.B.P., Santos F.R., Schneider M.P.C., Senavez H.N.,
 RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
 RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
 RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
 RA Vettore A., Wassen R., Zaha A., Simpson A.J.G.;
 RT "The complete genome sequence of Chromobacterium violaceum reveals
 RT remarkable and exploitable bacterial adaptability";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
 CC -!- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-oxobutanoate + H(2)O.
 CC -!- COFACTOR: Binds 1 4Fe-4S cluster (Potential).
 CC -!- PATHWAY: Valine and isoleucine biosynthesis; fourth step.
 CC -!- SIMILARITY: Belongs to the ilvd / edd family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE016914; AAQ58952.1; -.
 DR HAMAP; MF 00012; -; 1.
 DR PROSITE; PS00886; ILVD EDD 1; 1.
 DR PROSITE; PS00887; ILVD EDD 2; 1.
 KW Branched-chain amino acid biosynthesis; Lyase; Iron; Iron-sulfur;
 KW 4Fe-4S; Complete proteome.
 FT METAL 122
 FT METAL 195
 SQ SEQUENCE 618 AA; 65582 MW; 3273399ABAL92A05 CRC64;
 Query Match 34.2%; Score 44.5; DB 1; Length 618;
 Best Local Similarity 30.0%; Pred. No. 89;
 Matches 9; Conservative 7; Mismatches 7; Indels 7; Gaps 1;
 QY 3 GLPAVVGLSPGEQYHRCGVGL 25
 Db 307 GVPCLSKVAPATQKXHMEDVHRAGGVIGIL 336
 RESULT 15
 M18A HUMAN
 ID M18A HUMAN STANDARD; PRT; 2054 AA.
 AC Q92614; O81XP8;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Myosin XVIIIa (Myosin 18A) (Myosin containing PDZ domain).
 GN MYO18A OR MYSPDZ OR KIAA0216.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC MEDLINE=97191544; PubMed=9039502;
 RX Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawabayasi Y.,
 RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. VI.
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain.";
 RL DNA Res. 3:321-329(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Testis;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Colling F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Murray D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]

```

RX ALTERNATIVE SPLICING (ISOFORMS 1 AND 2).
RA MEDLINE=22646225; PubMed=12761286;
RA Mori K., Futusawa T., Okubo T., Inoue T., Ikawa S., Yanai N.,
RT "Genome structure and differential expression of two isoforms of a
RT novel PDZ-containing myosin (MysPDZ) (Mys18A).";
RL J. Biochem. 133:405-413 (2003).
CC -!- FUNCTION: May be involved in the maintenance of the stromal cell
CC architectures required for cell to cell contact (By similarity).
CC -!- SUBCELLULAR LOCATION: Isoform 1 co-localizes with the
CC endoplasmatic reticulum-Golgi complex; isoform 2, which lacks the
CC PDZ domain, is diffusely localized in the cytoplasm (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1; Synonyms=Alpha;
CC IsoId=Q92614-1; Sequence=Displayed;
CC Name=2; Synonyms=Beta;
CC IsoId=Q92614-2; Sequence=VSP_007869, VSP_007870;
CC Name=3;
CC IsoId=Q92614-3; Sequence=VSP_007871, VSP_007872;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -!- CAUTION: The TIAF1 protein is coded in the 3'UTR region of MysPDZ.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D86970; BAA13206.2; ALT INIT.
DR EMBL; BC039612; AAB39612.1; -.
DR PIR; PT0271; PT0271.
DR GO; GO:0005653; C:perinuclear space; TAS.
DR GO; GO:0006916; P:anti-apoptosis; TAS.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR003345; M_repeat.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF02370; M; 7.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR Myosin; ATP-binding; Coiled coil; Alternative splicing.
DR DOMAIN 220 311 PDZ.
DR DOMAIN 420 1186 MYOSIN HEAD-LIKE.
DR DOMAIN 1188 1217 IQ.
DR DOMAIN 1246 1971 COILED COIL (POTENTIAL).
DR NP_BIND 498 505 ATP (POTENTIAL).
DR VARSPPLIC 1 331 Missing (in isoform 2).
FT FTID=VSP_007869
FT SD -> MR (in isoform 2).
FT FTID=VSP_007870
FT Missing (in isoform 3).
FT FTID=VSP_007871
FT Missing (in isoform 3).
FT FTID=VSP_007872
FT SEQUENCE 2054 AA; 233113 MW; 52BFA0AA273E18F7 CRC64;
Query Match 34.2%; Score 44.5; DB 1; Length 2054;
Best Local Similarity 38.5%; Pred. No. 3.1e+02;

```

```

Matches 10; Conservative 3; Mismatches 4; Indels 9; Gaps 1;
QY 7 VVGLSPGEQF-----YHRCGVG 23
Db 651 VLGISPDEQKACWFIILAAIYHLGAAG 676
Search completed: May 7, 2004, 12:34:30
Job time : 11 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 46.75 seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-5

Perfect score: 130

Sequence: 1 MAGLPAVGLSPGEQYHRGGVGL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriophage.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67.5	51.9	41	11	Q8K408
2	56	43.1	242	16	Q89Y25
3	54	41.5	330	16	Q87YMS
4	52	40.0	150	16	Q8PE55
5	51	39.2	426	17	Q9YBR0
6	50.5	38.8	383	11	Q8K2L1
7	50.5	38.8	601	16	Q8FVA5
8	50.5	38.8	615	16	Q8VD21
9	50.5	38.8	662	11	Q8VHY3
10	50.5	38.8	680	11	Q8SSX1
11	50.5	38.8	758	10	Q8L525
12	50	38.5	277	16	Q9KQK6
13	50	38.5	284	16	Q7VFI5
14	50	38.5	335	16	Q82RR7
15	50	38.5	381	16	Q9RRL5
16	49.5	38.1	192	15	Q994Q2

17	49.5	38.1	192	15	Q9QSR2
18	49.5	38.1	603	16	Q8UAZ8
19	49	37.7	90	15	Q9QQA2
20	49	37.7	90	15	Q9QQA0
21	49	37.7	90	15	Q9QQA1
22	49	37.7	90	15	Q9QQA6
23	49	37.7	277	16	Q87RD3
24	49	37.7	331	16	Q7WEG7
25	49	37.7	331	16	Q7W350
26	49	37.7	331	16	Q7W032
27	49	37.7	363	13	Q9DDH0
28	49	37.7	711	15	Q7SVK4
29	49	37.7	711	15	Q7SVK3
30	49	37.7	713	15	Q7SVK2
31	49	37.7	753	16	Q89T31
32	49	37.7	814	16	Q8PHY9
33	49	37.7	814	16	Q8PE65
34	49	37.7	4823	13	Q93321
35	48.5	37.3	192	15	Q90CX5
36	48.5	37.3	192	15	Q91W51
37	48.5	37.3	192	15	Q8UTJ5
38	48.5	37.3	192	15	Q70887
39	48.5	37.3	195	3	Q9USC7
40	48	36.9	151	16	Q7U7G9
41	48	36.9	192	15	Q8UTQ8
42	48	36.9	192	15	Q90D29
43	48	36.9	271	16	Q87F13
44	48	36.9	278	16	Q9CMF6
45	48	36.9	300	16	Q8EMZ7

ALIGNMENTS

RESULT 1

Q8K408
ID Q8K408 PRELIMINARY; PRT; 41 AA.
AC Q8K408;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW.cl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E., Sood R., Hashimoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (CSF1) causes
osteopetrosis in the tl rat.";
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AA54137.1; -;
SQ SEQUENCE 41 AA; 4178 MW; 1D342C19BD18AA41 CRC64;

Query Match 51.9%; Score 67.5; DB 11; Length 41;
Best Local Similarity 68.0%; Pred. No. 0.017; 7; Indels 1; Gaps 1;

Matches 17; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MAGLPAVGLSPGEQYHRGGVGL 25

Db 18 MDGLPAAAGLSPEQCCCR-GVGL 41

RESULT 2

Q89Y25
ID Q89Y25 PRELIMINARY; PRT; 242 AA.
AC Q89Y25;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)

```

DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DE 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
GN BLO130 protein.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasanoto S., Watanabe A., Idesawa K., Iriuguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Teirukuchi H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AF005935; BAC45395.1; -.
DR InterPro; IPR005493; Methyltransf_6.
DR Pfam; PF03737; Methyltransf_6; 1.
KW Complete proteome.
SQ SEQUENCE 242 AA; 25216 MW; 588F9C0B9396414B CRC64;

Query Match 43.1%; Score 56; DB 16; Length 242;
Best Local Similarity 57.1%; Pred. No. 5.5;
Matches 12; Conservative 1; Mismatches 4; Indels 4; Gaps 1;

QY 1 MAGLPAVVGSLSPGQYHRGG 21
: ||| ||| ||| |||
Db 63 LARLEGAUGLKP----YHRGG 79

RESULT 3
Q87YM5 PRELIMINARY; PRT; 330 AA.
AC Q87YM5;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Conserved hypothetical protein.
GN PSPT03770.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., Deboy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidgen T.,
RA White O., Fraser C., Collier A.;
RT "Complete sequence of Pseudomonas syringae."
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016869; AAO57239.1; -.
DR TIGR; PSPT03770; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 330 AA; 36369 MW; 99876752E292FA71 CRC64;

Query Match 41.5%; Score 54; DB 16; Length 330;
Best Local Similarity 43.5%; Pred. No. 15;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 3 GLPAVVGSLSPGQYHRGGVGL 25
: ||| ||| ||| |||
Db 308 GQPLINGLAPSEAVFPGGIGKL 330

RESULT 4
Q8PE55 PRELIMINARY; PRT; 150 AA.
AC Q8PE55;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Deoxycytidylate deaminase.
GN XCC0126.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Parah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE012107; AAM39445.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR002125; dCMP_cyt deam.
DR Pfam; PF00383; dCMP_cyt deam; 1.
KW Complete proteome.
SQ SEQUENCE 150 AA; 16201 MW; 63D9C17D44DC8B43 CRC64;

Query Match 40.0%; Score 52; DB 16; Length 150;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGLPAVVGSLSPGQYH 18
||: ||| ||| ||| |||
Db 104 AGIKRWVALPGSESH 120

RESULT 5
Q9YBRO PRELIMINARY; PRT; 426 AA.
AC Q9YBRO;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein APE1539.
GN APE1539.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki C., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;

```

RT "Complete genome sequence of an aerobic hyper-thermophilic
RT Crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AF000061; BAA80538.1; -.
DR PIR; D72635; D72635.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 426 AA; 48461 MW; 6D3D3469392D4DF0 CRC64;

Query Match 39.2%; Score 51; DB 17; Length 426;
Best Local Similarity 45.5%; Pred. No. 54;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 3 GLPAVVGSLSPGEQYHRGGV 24
Db 278 GLADILGIEPGRREGRGV 299
|||:::|:|:|:|:|:|:|

RESULT 6

ID Q8K2L1 PRELIMINARY; PRT; 383 AA.
AC Q8K2L1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Procollagen, type IX, alpha 3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]__
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DDJ databases.
DR EMBL; BC030945; AAH30945.1; -.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 6.
DR ProDom; PD000007; Clg_helix; 3.
KW Collagen.
SQ SEQUENCE 383 AA; 35980 MW; B5F5AF37FF3192CE CRC64;

Query Match 38.8%; Score 50.5; DB 11; Length 383;
Best Local Similarity 52.0%; Pred. No. 57;
Matches 13; Conservative 3; Mismatches 6; Indels 3; Gaps 2;

QY 3 GLPAVVGSLSPGE--QYHRGGVGL 25
Db 66 GIPGDVG-V-PGEAGHRSYGAL 89
|:|:|:|:|:|:|:|:|:|

RESULT 7

Q8FVA5
ID Q8FVA5 PRELIMINARY; PRT; 601 AA.
AC Q8FVA5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Dehydratase, IlvD/Edd family.
GN BRA0940.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=2247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Ksaal B., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;

RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR EMBL; AE014587; AAN34112.1; -.
DR TIGR; BRA0940; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000581; ILVD_EDD_family.
DR Pfam; PF00920; ILVD_EDD; 1.
DR ProDom; PD002691; ILVD_EDD_family; 1.
DR PROSITE; PS00886; ILVD_EDD_1; 1.
KW Complete proteome.

QY 4 LPVAVGLSP-----GEQYHRGGV 22
Db 310 IPLVNLQPAGEYLGEDYHAGV 333
:|:|:|:|:|:|:|:|:|

Query Match

Best Local Similarity 38.8%; Score 50.5; DB 16; Length 601;
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 4 LPVAVGLSP-----GEQYHRGGV 22
Db 310 IPLVNLQPAGEYLGEDYHAGV 333
:|:|:|:|:|:|:|:|:|

RESULT 8

Q8YD21 PRELIMINARY; PRT; 615 AA.
ID Q8YD21
AC Q8YD21;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Galactonate dehydratase (EC 4.2.1.6).
GN BMEI10356.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mueh C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Ietesson J.-J.,
RA Haselkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009673; AAL53598.1; -.
DR PIR; AC3554; AC3554.
DR GO; GO:000869; F:galactonate dehydratase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000581; ILVD_EDD_family.
DR Pfam; PF00920; ILVD_EDD; 1.
DR ProDom; PD002691; ILVD_EDD_family; 1.
DR PROSITE; PS00886; ILVD_EDD_1; 1.
KW Lyase; Complete proteome.

SQ SEQUENCE 615 AA; 66750 MW; F7DC4B656C101438 CRC64;
Query Match 38.8%; Score 50.5; DB 16; Length 615;
Best Local Similarity 45.8%; Pred. No. 96;
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 4 LPVAVGLSP-----GEQYHRGGV 22
Db 324 IPLVNLQPAGEYLGEDYHAGV 347
:|:|:|:|:|:|:|:|:|

RESULT 9

Q8VHY3 PRELIMINARY; PRT; 662 AA.
ID Q8VHY3
AC Q8VHY3;

```

DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Alpha-3 type IX collagen (Fragment).
GN COL9A3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Swiss white; TISSUE=Urogenital ridge;
RA McClive P.J., Bell K.M., Sinclair A.H.;
RT "Type II and type IX collagen genes are expressed male-specifically in
RT the developing mouse gonad.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF349718; AAL56219.1; -.
DR MGD; MGI:894686; Col9a3.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 11.
DR ProDom; PD000007; Clg_helix; 3.
KW Collagen.
FT NON TER
FT SEQUENCE 662 AA; 61726 MW; 9261FFBFBF1AC4B7 CRC64;
SQ
Query Match 38.8%; Score 50.5; DB 11; Length 662;
Best Local Similarity 52.0%; Pred. No. 1.e+02;
Matches 13; Conservative 3; Mismatches 6; Indels 3; Gaps 2;

QY 3 GLPAVVGSLSPGE--QYHRGGVGVL 25
| : | | | : | | | : | | | |
Db 345 GLPGDVGV-PGERGAGRGSVGL 368

RESULT 10
Q8BSX1 PRELIMINARY; PRT; 680 AA.
AC Q8BSX1
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Procollagen.
GN COL9A3
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Xiphoid cartilage;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK030350; BAC28915.1; -.
DR MGD; MGI:894686; Col9a3.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 11.
DR ProDom; PD000007; Clg_helix; 2.
SQ SEQUENCE 680 AA; 63501 MW; 1939E44571032AE4 CRC64;

Query Match 38.8%; Score 50.5; DB 11; Length 680;
Best Local Similarity 52.0%; Pred. No. 1.e+02;
Matches 13; Conservative 3; Mismatches 6; Indels 3; Gaps 2;

QY 3 GLPAVVGSLSPGE--QYHRGGVGVL 25
| : | | | : | | | : | | | |
Db 363 GLPGDVGV-PGERGAGRGSVGL 386

RESULT 11
Q8L525 PRELIMINARY; PRT; 758 AA.
AC Q8L525
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE B1103C09.11 protein (P0451D05.19 protein).
GN B1103C09.11 OR P0451D05.19.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriatoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
RT clone:B1103C09.11";
DR EMBL; AP003333; BAB91821.1; -.
DR EMBL; AP003253; BAB92317.1; -.
DR Gramene; Q8L525; -.
DR SEQUENCE 758 AA; 81761 MW; 2A95E763198FE12C CRC64;
SQ
Query Match 38.8%; Score 50.5; DB 10; Length 758;
Best Local Similarity 32.4%; Pred. No. 1.2e+02;
Matches 12; Conservative 3; Mismatches 7; Indels 15; Gaps 1;

QY 3 GLPAVVGSLSPGEY-----HRGGVG 24
| : | | | : | | | : | | | |
Db 552 GAAGAIGLAQGEQNFPGTPTALLPVMGFSGQHPGGVG 588

RESULT 12
Q9KQK6 PRELIMINARY; PRT; 277 AA.
AC Q9KQK6
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Formyltetrahydrofolate deformylase.
GN VCI192.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
DR EMBL; AE004274; AAF95140.1; -.
DR FIC; F82130; F82130.
DR HSSP; P08179; 2GAR.

```

DR TIGR; VC1992; --
 DR GO; GO:0016537; F:amino acid binding; IEA.
 DR GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
 DR GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. .; IEA.
 DR GO; GO:0006189; P:de novo IMP biosynthesis; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR002376; formyl_transf.
 DR InterPro; IPR004810; PurU.
 DR Pfam; PF01842; ACT; 1.
 DR Pfam; PF00551; formyl_transf; 1.
 DR PRINTS; PR01575; FFH4HYDRLASE.
 DR TIGRFAMS; TIGR00655; PurU; 1.
 DR Complete proteome.
 SQ SEQUENCE 277 AA; 31373 MW; A703491654753DC6 CRC64;
 Query Match 38.5%; Score 50; DB 16; Length 277;
 Best Local Similarity 52.9%; Pred. No. 47;
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 Qy 4 LPVAVGLSPGGEQYHRG 20
 ||||| :||| :|||
 Db 190 LPAFIGAGPYQQAVERG 206
 ||||| :||| :|||
 RESULT 13
 Q7VF15 PRELIMINARY; PRT; 284 AA.
 AC Q7VF15;
 DT 01-OCT-2003 (TRENBLrel. 25, Created)
 DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Formyltetrahydrofolate deformylase PurU (EC 3.5.1.10).
 GN PURU OR RH1691.
 OS Helicobacter hepaticus.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=32025;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 51449 / 3B1;
 RX MEDLINE=22709201; PubMed=12810954;
 RA Suerbaum S., Josenhans C., Sterzenbach T., Drescher B., Brandt P.,
 Bell M., Droge M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
 Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
 Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
 RA "The complete genome sequence of the carcinogenic bacterium
 RT Helicobacter hepaticus".
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906 (2003).
 DR EMBL; AB017149; AAF78288.1; --
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 284 AA; 32660 MW; AAED2DC5086236E3 CRC64;
 Query Match 38.5%; Score 50; DB 16; Length 284;
 Best Local Similarity 52.9%; Pred. No. 49;
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 Qy 4 LPVAVGLSPGGEQYHRG 20
 ||||| :||| :|||
 Db 197 LPAFIGANPYQQAVERG 213
 ||||| :||| :|||
 RESULT 14
 Q82RR7 PRELIMINARY; PRT; 335 AA.
 AC Q82RR7;
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Putative terpene cyclase.
 GN TPCL1 OR SAV76.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RA "Genome sequence of an industrial microorganism Streptomycetes
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites".
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 Sakaki Y., Hattori M., Omura S.;
 RA "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomycetes avermitilis".
 RL Nat. Biotechnol. 21:526-531 (2003).
 DR EMBL; AP005021; BAC67785.1; --
 DR InterPro; IPR008949; Terpenoid_synth.
 DR Complete proteome.
 SQ SEQUENCE 335 AA; 36480 MW; 49B8477E2D52666F CRC64;
 Query Match 38.5%; Score 50; DB 16; Length 335;
 Best Local Similarity 47.6%; Pred. No. 58;
 Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
 Qy 3 GLPAVGLSPGGEQYHRGGVG 23
 ||||| :||| :|||
 Db 8 GLPAPAGISPGLEATRRHNLG 28
 ||||| :||| :|||
 RESULT 15
 Q9RR15 PRELIMINARY; PRT; 381 AA.
 AC Q9RR15;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
 DE Cytochrome P450.
 GN DR2473.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heideberg J.F., Hickey E.K., Peterson J.D.,
 Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 Vanatavan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
 Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 Fraser C.M.;
 RA "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans R1".
 RL Science 286:1571-1577 (1999).
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AE002076; AAF12016.1; --
 DR PIR; F75270; F75270.
 DR TIGR; DR2473; --
 DR GO; GO:0004937; F:monooxygenase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.

KW Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ SEQUENCE 381 AA; 41940 MW; F191EA69F1797B53 CRC64;

Query Match 38.5%; Score 50; DB 16; Length 381;
Best Local Similarity 100.0%; Pred. NO. 67;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GLPAVVGLSP 12
|||
Db 51 GLPAVVGLSP 60

Search completed: May 7, 2004, 12:37:50
Job time : 48.9167 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 69.75 seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-5

Perfect score: 130

Sequence: 1 MAGLPVAVGLSPGEQYHRGGVGL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_20Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	130	100.0	25	3	AAY84264
2	106	81.5	20	3	AAY84265
3	75	57.7	15	3	AAY84269
4	72	55.4	14	3	AAY84266
5	68	52.3	13	3	AAY84267
6	65	50.0	13	3	AAY84268
7	53	40.8	394	4	AAU43816
8	53	40.8	394	6	ABM40335
9	52	40.0	187	6	ABU21296
10	51.5	39.6	191	3	AAB69296
11	50	38.5	277	6	ABU49441
12	49.5	38.1	192	3	AAB69298
13	49	37.7	278	6	ABU30626
14	48.5	37.3	192	3	AAB69290
15	48	36.9	278	6	ABU39195
16	48	36.9	287	6	ABU17217
17	48	36.9	303	6	ADA36829
18	47.5	36.5	114	4	AAE03295
19	47.5	36.5	114	5	ABG64478
20	47	36.2	151	5	ABU50765
21	47	36.2	155	5	ABU51750
22	47	36.2	238	6	ABU27483
23	47	36.2	278	6	ABU44771
24	47	36.2	280	6	ABU48054
25	47	36.2	280	6	ABU31449

26	47	36.2	280	6	ABU15054
27	47	36.2	280	6	ABU47098
28	47	36.2	282	6	ABU41187
29	47	36.2	283	2	ABM69037
30	47	36.2	293	2	AAW98485
31	47	36.2	293	6	ABU31073
32	46.5	35.8	656	6	ABU27276
33	46	35.4	46	3	AAV58745
34	46	35.4	143	5	AAO21677
35	46	35.4	143	7	ADB64633
36	46	35.4	184	5	ABE26551
37	46	35.4	234	4	ABE36208
38	46	35.4	274	6	ABU26498
39	46	35.4	287	7	ADB80057
40	46	35.4	300	2	AAV16108
41	46	35.4	418	5	ABE04480
42	46	35.4	450	5	ABH04479
43	46	35.4	692	6	ABR40877
44	46	35.4	921	2	AAV08304
45	46	35.4	1070	7	ADE08073

ALIGNMENTS

RESULT 1

AAV84264

ID AAY84264 standard; peptide; 25 AA.

XX AC AAY84264;

XX DT 12-JUL-2000 (first entry)

XX DE Peptide of alternate reading frame of macrophage colony stimulating gene.

XX KW Renal cell carcinoma; antigen; cytotoxic T lymphocyte;

XX KW tumour rejection antigen; macrophage colony stimulating gene;

XX KW macrophage-colony stimulating factor; antigen presenting cell;

XX KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX OS Homo sapiens.

XX PN WO200013699-A1.

XX PD 16-MAR-2000.

XX PF 03-SEP-1999; 99WO-US020344.

XX PR 04-SEP-1998; 98US-0099077P.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX PS WPI; 2000-256859/22.

XX PT N-PSDB; AAZ99672.

XX PT Isolated polypeptide used to treat subjects having a disorder

XX PT characterized by expression of alternative open reading frame macrophage-

XX PT colony stimulating factor comprises 25 amino acid residue sequence.

XX PS Claim 1; Page 64; 74pp; English.

XX CC The present sequence represents a tumour rejection antigen precursor, and

XX CC is encoded by an alternative open reading frame (ORF) of human macrophage

XX CC colony stimulating gene. Peptides derived from the alternative ORF of

XX CC macrophage-colony stimulating factor, when presented by an antigen

XX CC presenting cell having a human leukocyte antigen (HLA) class I molecule,

XX CC effectively induce the activation and proliferation of CD8+ cytotoxic T

XX CC lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF

XX CC of macrophage-colony stimulating factor are useful for enriching

XX CC selectively a population of T lymphocytes with CD8+ T lymphocytes. They

XX CC are also useful for diagnosing a disorder characterized by expression of

CC the polypeptide, and for identifying functional variants and mimetics
 XX
 SQ Sequence 25 AA;

Query Match 100.0%; Score 130; DB 3; Length 25;
 Best Local Similarity 100.0%; Pred. No. 2.3e-11;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGLPAVVGSLSPGEQYHRGGVGL 25
 |||||
 Db 1 MAGLPAVVGSLSPGEQYHRGGVGL 25

RESULT 2
 AAY84265
 ID AAY84265 standard; peptide; 20 AA.

XX
 AC AAY84265;
 XX
 DT 12-JUL-2000 (first entry)
 XX
 DE Truncated macrophage colony stimulating factor tumour antigen.
 XX
 KW tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX
 OS Homo sapiens.
 XX
 PN WO200013699-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 03-SEP-1999; 99WO-US020344.
 XX
 PR 04-SEP-1998; 98US-0099077P.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.

XX
 PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
 XX
 DR WPI: 2000-256859/22.
 DR N-PSDB; AAZ99675.
 XX
 PT Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Claim 3; Page 64; 74pp; English.

XX The present sequence represents a truncated tumour rejection antigen
 CC precursor, and is encoded by a truncated alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 20 AA;

Query Match 81.5%; Score 106; DB 3; Length 20;
 Best Local Similarity 100.0%; Pred. No. 4.7e-08;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGLPAVVGSLSPGEQYHRG 20
 |||||
 Db 1 MAGLPAVVGSLSPGEQYHRG 20

RESULT 3
 AAY84269

ID AAY84269 standard; peptide; 15 AA.

XX
 AC AAY84269;
 XX
 DT 12-JUL-2000 (first entry)

XX Peptide derived from macrophage colony stimulating gene alternative ORF.
 DE
 XX
 KW tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN WO200013699-A1.

XX
 PD 16-MAR-2000.
 XX
 PF 03-SEP-1999; 99WO-US020344.

XX
 PR 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI: 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 15 AA;

Query Match 57.7%; Score 75; DB 3; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.00087;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGLPAVVGSLSPGEQ 16
 |||||
 Db 1 AGLPAVVGSLSPGEQ 15

RESULT 4

AAY84266

ID AAY84266 standard; peptide; 14 AA.

XX

AC AAY84266;

DT 12-JUL-2000 (first entry)

XX Peptide derived from macrophage colony stimulating gene alternative ORF.
 DE
 XX
 KW tumour rejection antigen; macrophage colony stimulating gene;

KW macrophage-colony stimulating factor; antigen presenting cell;
 XX human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

OS Synthetic.
 OS Homo sapiens.
 XX WO200013699-A1.
 XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.
 XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
 XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder

PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.
 XX Claim 2; Page 39; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour

CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 14 AA;

Query Match 55.4%; Score 72; DB 3; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.0022;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LPAVVGSLSPGQEQY 17
 Db 1 LPAVVGSLSPGQEQY 14

RESULT 5
 AAY84267
 ID AAY84267 standard; peptide; 13 AA.

AC AAY84267;

XX 12-JUL-2000 (first entry)

XX Peptide derived from macrophage colony stimulating gene alternative ORF.

XX tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Synthetic.
 OS Homo sapiens.

XX WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX

PR 04-SEP-1998; 98US-0099077P.
 XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder

PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 13 AA;

Query Match 52.3%; Score 68; DB 3; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.0074;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PAVVGSLSPGQEQY 17
 Db 1 PAVVGSLSPGQEQY 13

RESULT 6

AAY84268
 ID AAY84268 standard; peptide; 13 AA.

AC AAY84268;

XX 12-JUL-2000 (first entry)

XX Peptide derived from macrophage colony stimulating gene alternative ORF.

XX tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Synthetic.
 OS Homo sapiens.

XX WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder

PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX

PS Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics

XX Sequence 13 AA;

Query Match 50.0%; Score 65; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LPVVGLSPGEQE 16
DB 1 LPVVGLSPGEQE 13

RESULT 7

AAU43816
ID AAU43816 standard; protein; 394 AA.

XX AC AAU43816;

XX 13-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #4712.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208941P.

XX 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59521.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.

XX Claim 6; SEQ ID NO 5011; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX Sequence 394 AA;

Query Match 40.8%; Score 53; DB 4; Length 394;
Best Local Similarity 45.0%; Pred. No. 39;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 5 PAVVGLSPGEQEHYHGGVGV 24
DB 255 PVVLGTAPGQGHDRHGTGV 274

RESULT 8

ABM40335

ID ABM40335 standard; protein; 394 AA.

XX AC ABM40335;

XX 20-OCT-2003 (first entry)

XX Propionibacterium acnes immunogenic polypeptide #5011.

XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine; immunogenic.

XX Propionibacterium acnes.

XX WO2003033515-A1.

XX 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.

XX 15-OCT-2001; 2001US-00978825.

XX (CORI-) CORIXA CORP.

XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieve-Douglass J;

XX WPI; 2003-381789/36.

XX N-PSDB; ACF64450.

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.

XX Claim 6; SEQ ID NO 5011; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC polynucleotide of the invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared

via this method; a vaccine composition (comprising P. acnes polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations, or antigen-presenting cells that express the polypeptide); a method and kit for detecting or determining the presence or absence of P. acnes in a patient; and a method for inhibiting the development of P. acnes in a patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations or antigen-presenting cells that express the polypeptides are useful for diagnosing, preventing or treating acne vulgaris, or for stimulating an immune response specific for a P. acnes protein. The polynucleotides can also be used as probes or primers for nucleic acid hybridisation. The vaccine composition is useful for the stimulation of an immune response against P. acnes, or for treating acne, and the kit is useful for performing a diagnostic assay. The present sequence represents a specifically claimed P. acnes polypeptide which is thought to contain an immunogenic region. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 40.8%; Score 53; DB 6; Length 394;
Best Local Similarity 45.0%; Pred. No. 39;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PAVVGLSPGEOYHRGGVGV 24
Db 255 PVVLTAPGGQHRHGTGV 274

Sequence 394 AA;

RESULT 9
ABU21296
ID ABU21296 standard; protein; 187 AA.
XX ABU21296;
AC ABU21296;
DT 19-JUN-2003 (first entry)
DE Protein encoded by Prokaryotic essential gene #6823.
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS Burkholderia fungorum.
PN WO200277183-A2.
XX 03-OCT-2002.
PD 21-MAR-2002; 2002WO-US009107.
PF 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA25166.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 49220; 1766pp; English.
PS The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:

(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than S. aureus, S. typhimurium, K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 187 AA;

Query Match 40.0%; Score 52; DB 6; Length 187;
Best Local Similarity 47.8%; Pred. No. 24;
Matches 11; Conservative 2; Mismatches 6; Indels 4; Gaps 1;

Qy 3 GLPA----VVLSPGEOYHRGG 21
Db 48 GLPVVLNVVGMFGRTEQHAG 70

RESULT 10
AAB69296
ID AAB69296 standard; protein; 191 AA.
XX AAB69296;
XX 12-SEP-2003 (revised)
DT 20-APR-2001 (first entry)
DE HIV-1 non-subtype B clone 94CY032-3 vif protein.
XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpr; vif; vpr; tat; rev; nef; vaccine.
XX Human immunodeficiency virus 1.
OS Human immunodeficiency virus 1.
XX WO200026416-A1.
XX 11-MAY-2000.
XX 25-OCT-1999; 99WO-US024837.
XX 02-NOV-1998; 98US-00184418.
XX (UABR-) UAB RES FOUND.
PA Hahn BH, Shaw GM, Gao F;
PI WPI; 2000-365651/31.
DR Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
XX type 1 useful for detecting and treating AIDS comprises a specific
PT nucleotide sequence.

XX Claim 41; Fig 16; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 191 AA;

Query Match 39.6%; Score 51.5; DB 3; Length 191;

Best Local Similarity 58.8%; Pred. No. 29;

Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 9 GLSPGEQYHRG-GVGV 24

Db 70 GLPGEQDHLGHGWSI 86

RESULT 11

ABU49441

ID ABU49441 standard; protein; 277 AA.

XX ABU49441;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #34968.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Vibrio cholerae.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX N-PSDB; ACA53311.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 77365; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 277 AA;

Query Match 38.5%; Score 50; DB 6; Length 277;

Best Local Similarity 52.9%; Pred. No. 71;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 LPVVGLSPGEQYHRG 20

Db 190 LPAFIGAKPYQQAYERG 206

RESULT 12

AAB69298

ID AAB69298 standard; protein; 192 AA.

XX AAB69298;

XX 12-SEP-2003 (revised)

XX 20-APR-2001 (first entry)

XX HIV-1 non-subtype B clone 96ZM751-3 vif protein.

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;

XX vif; vpr; tat; rev; nef; vaccine.

XX Human immunodeficiency virus 1.

XX WO200026416-A1.

XX 11-MAY-2000.

XX 25-OCT-1999; 99WO-US024837.

XX 02-NOV-1998; 98US-00184418.

XX (UABR-) UAB RES FOUND.

XX Hahn BH, Shaw GM, Gao F;

XX WPI; 2000-365651/31.

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific

PT nucleotide sequence.

XX Claim 41; Fig 16; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,


```

CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 278 AA;
Query Match 36.9%; Score 48; DB 6; Length 278;
Best Local Similarity 52.9%; Pred. No. 1.4e+02;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0
QY 4 LPAVVGSLSPGQEVYHRG 20
||| : |||
Db 191 LPAFIGAKPYHQAYERG 207
||| : |||
Search completed: May 7, 2004, 12:33:42
Job time : 71.75 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 7.2 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-9

Perfect score: 106

Sequence: 1 MAGLPAVVGSLSPGQEXHRG 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	49	46.2	278	1	PURU_HAEIN
2	48	45.3	780	1	PFSA_DEIRA
3	47	44.3	280	1	PURU_ECOL6
4	47	44.3	280	1	PURU_ECOL1
5	46	43.4	508	1	COBI_MYCTU
6	44	41.5	1172	1	CNA2_MOUSE
7	44	41.5	3664	1	MINT_HUMAN
8	43	40.6	211	1	COBL_METUA
9	43	40.6	326	1	COBL_MYCTU
10	43	40.6	390	1	COBL_MYCTU
11	43	40.6	505	1	Y76J_CAEEL
12	43	40.6	516	1	C4AD_DROME
13	43	40.6	544	1	CH60_PROAC
14	43	40.6	628	1	HNFA_MOUSE
15	43	40.6	628	1	HNFA_RAT
16	43	40.6	631	1	HNFA_HUMAN
17	43	40.6	633	1	PLB5_SCHPO
18	43	40.6	658	1	VG18_BPT4
19	43	40.6	677	1	SG1_HUMAN
20	43	40.6	983	1	RPOB_SUNNV
21	42	39.6	400	1	TRUD_METRA
22	42	39.6	469	1	GLXJ_YEAST
23	42	39.6	808	1	PLSB_VIBPA
24	42	39.6	876	1	TOP1_VIBCH
25	42	39.6	1065	1	RPOB_WARPO
26	42	39.6	1712	1	CA24_HUMAN
27	41	38.7	156	1	RVUX_CAUCR
28	41	38.7	213	1	MDCG_XANAC
29	41	38.7	243	1	SUMT_SNP7
30	41	38.7	479	1	PTSE_VIBAL
31	41	38.7	481	1	GYRB_FLEJA
32	41	38.7	520	1	AT15_YEAST
33	41	38.7	579	1	NH71_CAEEL

34	41	38.7	774	1	LOL2_HUMAN
35	41	38.7	934	1	OD01_COXBU
36	41	38.7	981	1	RRPO_AHNNV
37	40.5	38.2	387	1	OMA2_NEIMC
38	40.5	38.2	392	1	OMA2_NEIMC
39	40.5	38.2	393	1	OMA1_NEIMC
40	40.5	38.2	395	1	OMA1_NEIMA
41	40.5	38.2	811	1	PRI3_RHOFU
42	40	37.7	157	1	RISB_PYRFU
43	40	37.7	193	1	VIF_SIVCZ
44	40	37.7	316	1	Y273_SYNY3
45	40	37.7	331	1	KDD2_STRAW

ALIGNMENTS

RESULT 1
PURU_HAEIN STANDARD; PRT; 278 AA.
AC Q03432;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-PH(4) hydrolase).
GN PURU OR H11588.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=9530630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Cotton M.D., Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512 (1995).
RN [2]
RP SEQUENCE OF 64-278 FROM N.A.
RC STRAIN-RM 7004 / Serotype B;
RX MEDLINE=93328119; PubMed=8335255;
RA Maskell D.J.;
RT "Cloning and sequencing of the Haemophilus influenzae *aroA* gene.";
RL Gene 129:155-156 (1993).
CC -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate + tetrahydrofolate.
CC -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine (By similarity).
CC -!- PATHWAY: De novo purine biosynthesis.
CC -!- SUBUNIT: Homohexamer (By similarity).
CC -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32833; AAC23236.1; -.
DR EMBL; L04686; AAA24942.1; -.
DR PIR; E64131; E64131.
DR HSSP; P08179; IGRC.
DR TIGR; H11588; -.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR002376; formyl_transf.
DR InterPro; IPR004810; PurU.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00551; formyl_transf; 1.
DR PRINTS; PR01575; FFH4HYDRLASE.
DR TIGRFAMs; TIGR00655; PurU; 1.
KW Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW Complete proteome.
FT ACT_SITE 223 BY SIMILARITY.
FT CONFLICT 115 VIG -> RNR (IN REF. 2).
FT CONFLICT 138 HEN -> PK (IN REF. 2).
FT CONFLICT 205 K -> E (IN REF. 2).
SQ SEQUENCE 278 AA; 32173 MW; 7F375AB3C422EC4B CRC64;

Query Match 46.2%; Score 49; DB 1; Length 278;
Best Local Similarity 52.9%; Pred. No. 3.6;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 LPAAVGLSPGCEQYVHRG 20
DB 191 LPAAFGAKPYQAYKRG 207

RESULT 2
PPSA DEIRA
ID PPSA DEIRA STANDARD; PRT; 780 AA.
AC O83026;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phosphoenolpyruvate synthase (EC 2.7.9.2) (Pyruvate, water dikinase)
DE (PEP synthase).
GN PPSA OR DRI1727.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
EX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pauphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
RN [2]
RP SEQUENCE OF 259-780 FROM N.A.
RA Narumi I., Islam S., Cherdchu K., Kikuchi M., Watanabe H.,
RA Kitayama S., Yamamoto K.;
RT "IS8301: the second insertion sequence element from Deinococcus
RT radiodurans.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: ATP + pyruvate + H(2)O = AMP +
CC phosphoenolpyruvate + phosphate.
CC -|- PATHWAY: ESSENTIAL STEP IN GLUCONOGENESIS WHEN PYRUVATE AND
CC LACTATE ARE USED AS A CARBON SOURCE.
CC -|- SIMILARITY: Belongs to the PEP-utilizing enzyme family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002014; AAF11283.1; -.
DR EMBL; AB016803; BAA32387.1; -.
DR PIR; D75361; D75361.
DR PIR; T44369; T44369.
DR TIGR; DRI1727; -.
DR InterPro; IPR008279; PEP mobile.
DR InterPro; IPR006318; PEP_P trans.
DR InterPro; IPR006319; PEP synth.
DR InterPro; IPR000121; PEP utilizers.
DR InterPro; IPR002192; PPD_K term.
DR Pfam; PF00391; PEP-utilizers; 1.
DR Pfam; PF02896; PEP-utilizers; C; 1.
DR Pfam; PF01326; PPD_K term; 1.
DR PRINTS; PR01736; PHPTFRNFRASE.
DR PRODOM; PD000940; PEP utilizers; 1.
DR TIGRFAMs; TIGR01418; PEP synth; 1.
DR PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.
DR PROSITE; PS00742; PEP_ENZYMES_2; 1.
KW Transferase; Kinase; ATP-binding; Phosphorylation; Complete proteome.
FT ACT_SITE 409 409 TELE-PHOSPHOHISTIDINE INTERMEDIATE
FT ACT_SITE 409 409 (BY SIMILARITY).
FT MOD_RES 409 409 PHOSPHORYLATION (BY SIMILARITY).
FT SEQUENCE 780 AA; 84895 MW; AD555076324ADA7 CRC64;

Query Match 45.3%; Score 48; DB 1; Length 780;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 GLPAVVGSLSPGCEQYVHRG 20
DB 418 GLPAVVGSGNATRELHNG 435

RESULT 3
PURA ECOL6
ID PURA ECOL6 STANDARD; PRT; 280 AA.
AC P38480;
DT 01-OCT-1994 (Rel. 30, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
DE hydrolase).
GN PURA OR C1696 OR SF1232 OR S1318.
OS Escherichia coli O6, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=93023838; PubMed=1406252;
RA Homocckj A.E., Tucker S.C., Maurelli A.T.;
RT "Temperature regulation of Shigella virulence: identification of the
RT repressor gene virR, an analogue of hns, and partial complementation

```

RT by tyrosyl transfer RNA (tRNA^{Tyr}).";
 RL Mol. Microbiol. 6:2113-2124(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Ou D., Dong J.,
 RA Sun L., Xue X., Zhao X., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 RT through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441 (2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=2457T / AFCC 700930 / Serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Vance L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.;
 RT "Complete genome sequence and comparative genomics of Shigella
 RT flexneri serotype 2a strain 2457T.";
 RL Infect. Immun. 71:2775-2786(2003).
 RN [5]
 RP IDENTIFICATION.
 RX MEDLINE=94042872; PubMed=8226647;
 RA Nagy P.L., McCorkle G., Zalkin H.;
 RT "puru, a source of formate for purT-dependent phosphoribosyl-N-
 RT formylglycinamide synthesis.";
 RL J. Bacteriol. 175:7066-7073(1993).
 CC -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
 CC THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
 CC 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
 CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.
 CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
 CC tetrahydrofolate.
 CC -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine
 CC (by similarity).
 CC -!- PATHWAY: De novo purine biosynthesis.
 CC -!- SUBUNIT: Homohexamer (by similarity).
 CC -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE016760; AAN80163.1; -;
 DR EMBL; X66849; -; NOT ANNOTATED_CDS.
 DR EMBL; AE015150; AAN42845.1; -;
 DR EMBL; AE016982; AAP16730.1; -;
 DR HSSP; P08179; 1GPC
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR002376; formyl_transf.
 DR InterPro; IPR004810; PurU.
 DR Pfam; PF01842; ACT; 1.
 DR Pfam; PF00551; formyl_transf; 1.
 DR PRINTS; PR01575; FFH4HYDRLASE.
 DR TIGRPFAMs; TIGR00655; PurU; 1.
 KW Purine biosynthesis; Hydrolyase; One-carbon metabolism;
 KW Complete proteome.
 FT ACT SIZE 225 BY SIMILARITY.
 FT CONFLICT 44 R -> L (IN REF. 2).
 FT SEQUENCE 280 AA; 31920 MW; 55BC16B62727A419 CRC64;
 SQ
 Query Match 44.3%; Score 47; DB 1; Length 280;
 Best Local Similarity 52.9%; Pred. No. 7.4;
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 QY 4 LPVVVGLSPGEQYHRG 20
 DB 193 LPAFIGARPYHQAYERG 209
 RESULT 4
 PURU_ECOLI
 ID PURU_ECOLI STANDARD; PRT; 280 AA.
 AC P37051;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
 DE hydrolase).
 GN PURU OR TGS OR B1232.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-6.
 RC STRAIN=K12;
 RX MEDLINE=94042872; PubMed=8226647;
 RA Nagy P.L., McCorkle G., Zalkin H.;
 RT "puru, a source of formate for purT-dependent phosphoribosyl-N-
 RT formylglycinamide synthesis.";
 RL J. Bacteriol. 175:7066-7073(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94110230; PubMed=8282700;
 RA Boesl M., Kersten H.;
 RT "Organization and functions of genes in the upstream region of tyrt
 RT of Escherichia coli: phenotypes of mutants with partial deletion of a
 RT new gene [tgs].";
 RL J. Bacteriol. 176:221-231(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=95173107; PubMed=7868604;
 RA Nagy P.L., Marolewski A., Benkovic S.J., Zalkin H.;
 RT "Formyltetrahydrofolate hydrolase, a regulatory enzyme that functions
 RT to balance pools of tetrahydrofolate and one-carbon tetrahydrofolate
 RT adducts in Escherichia coli.";
 RL J. Bacteriol. 177:1292-1298(1995).
 CC -!- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
 CC THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
 CC 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
 CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.
 CC -!- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +

```

CC tetrahydrofolate.
CC -!- ENZYME REGULATION: Activated by methionine, inhibited by glycine.
CC -!- PATHWAY: De novo purine biosynthesis.
CC -!- SUBUNIT: Homohexamer.
CC -!- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L20251; AAC36846.1; -.
DR EMBL; M64675; AAA16860.1; ALT_INIT.
DR EMBL; AE000221; AAC74314.1; -.
DR EMBL; D90758; BAA36100.1; -.
DR EMBL; D90759; BAA36112.1; -.
DR EMBL; D90852; BAA16026.1; -.
DR PIR; C36871; C36871.
DR EcoGene; EG11819; purU.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR002376; formyl_transf.
DR InterPro; IPR004810; PurU.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00551; formyl_transf; 1.
DR PRINTS; PR01575; FFH4HYDRLASE.
DR TIGRFAMs; TIGR00655; PurU; 1.
KW Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW Complete proteome.
FT ACT_SITE 225 BY SIMILARITY.
SQ SEQUENCE 280 AA; 31934 MW; 5667406D2727A2C2 CRC64;

Query Match 44.3%; Score 47; DB 1; Length 280;
Best Local Similarity 52.9%; Pred. No. 7.4;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 LPVVVGLSPGQEQYVRG 20
DB 193 LPAFIGARPYHQVYERG 209

RESULT 5
ID COBI MYCTU STANDARD; PRT; 508 AA.
AC Q10677;
DT 15-JUL-1998 (Rel. 36, Created)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
DE methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine-precorrin-2
DE methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)]].
GN COBIJ OR COBI OR RV2066 OR MT2126 OR MTCV49.05 OR MB2092.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RN SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
```

```

RL Nature 393:537-544 (1998).
RN [2]
RN SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kelownay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RN SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC PRECORRIN-3A (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC adenosyl-L-homocysteine + precorrin-3A.
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC AND CBIL/COBI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z73966; CAA98214.1; -.
DR EMBL; AE007083; AAK46406.1; -.
DR EMBL; BX248341; CAD96945.1; -.
DR PIR; E70764; E70764.
DR TIGR; MT2126; -.
DR TuberculList; RV2066; -.
DR InterPro; IPR006364; Cobi_CbIL.
DR InterPro; IPR008363; CobiJ.
DR InterPro; IPR008078; Cor/por_Metransf.
DR InterPro; IPR003043; Uropor_Metransf.
DR Pfam; PF00590; TP_methylase; 2.
DR TIGRFAMs; TIGR01467; cobi_cbil; 1.
DR TIGRFAMs; TIGR01466; cobi_cbil; 1.
DR PROSITE; PS00839; SUMT_1; 1.
DR PROSITE; PS00840; SUMT_2; 1.
DR Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
DR Methyltransferase; Multifunctional enzyme; Complete proteome.
FT DOMAIN 1 243 PRECORRIN-2 C20-METHYLTRANSFERASE.
FT DOMAIN 244 508 PRECORRIN-3 METHYLASE.
SQ SEQUENCE 508 AA; 53910 MW; 95AC066F022C4DC1 CRC64;

Query Match 43.4%; Score 46; DB 1; Length 508;
Best Local Similarity 47.1%; Pred. No. 19;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAGLPVVVGLSPGQEQY 17
DB 245 LTGVVVVGLGPGSDW 261

RESULT 6
CNA2_MOUSE
ID CNA2_MOUSE STANDARD; PRT; 1172 AA.
AC Q80XC6; Q8R3D7; Q99LT9;
```

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein C14orf102 homolog.
GN C14orf102.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN CONCEPTUAL TRANSLATION, AND RECONSTRUCTION FROM EST.
RA Reynaud S.;
RL Unpublished observations (AUG-2003).
RN [2]
RN SEQUENCE OF 21-1172 FROM N.A.
RC TISSUE=Breast tumor, and Olfactory epithelium;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Helel F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- SIMILARITY: Contains 5 HAT repeats.
CC -!- CAUTION: This is a conceptual translation; a frameshift was
introduced in position 28 to produce the correct N-terminus, and
to extend the similarity with the human ortholog.
CC -!- CAUTION: Ref.1 (AAH51175) sequence differs from that shown due to
a stop codon in position 1146 which was translated as Gln to
extend the sequence.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; BY182441; -; NOT ANNOTATED_CDS.
DR EMBL; BC002230; AAH02230.1; -;
DR EMBL; BC025577; AAH25577.1; ALT INIT.
DR EMBL; BC051175; AAH51175.1; ALT_SEQ.
DR InterPro; IPR003107; HAT.
DR SMART; SM00386; HAT; 5.
KW Repeat; Coiled coil.
FT DOMAIN 78 330 COILED COIL (POTENTIAL).
FT REPEAT 314 346 HAT 1.
FT REPEAT 404 436 HAT 2.
FT REPEAT 766 800 HAT 3.
FT REPEAT 986 1018 HAT 4.
FT REPEAT 1075 1109 HAT 5.
FT DOMAIN 86 93 POLY-LYS.
FT DOMAIN 146 154 POLY-ALA.
FT DOMAIN 707 707 R -> H (IN REF. 1; AAH25577).
FT CONFLICT 741 742 LS -> VC (IN REF. 1; AAH25577).
SQ SEQUENCE 1172 AA; 133424 MW; ABE37A8BCE7A528F CRC64;
Query Match 41.5%; Score 44; DB 1; Length 1172;

Best Local Similarity 53.3%; Pred. No. 94;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 5 PAVVGLSPGQEVYHR 19
DB 351 PGYALGEGGEQEKHR 365
RESULT 7
MINT HUMAN
ID MINT_HUMAN STANDARD; PRT: 3664 AA.
AC Q96T58; Q9H9A8; Q9NWH5; Q9UQ01; Q9Y556;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIAA0929.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,
RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MED3;
RNA AND MTA1L1.
RC TISSUE=Liver, and Pituitary;
RX MEDLINE=21231190; PubMed=11331609;
RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
RA Hon M., Evans R.M.;
RT "Sharp, an inducible cofactor that integrates nuclear receptor
repression and activation.";
RL Genes Dev. 15:1140-1151 (2001).
RN [2]
RN SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 294-3664 FROM N.A.
RA Rhodes S., Huckle E.;
RN Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
RC TISSUE=Embryo, and Teratocarcinoma;
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE OF 2002-3664 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=992465063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 6:63-70 (1999).
RN [6]
RN INTERACTION WITH PPARD.
RP MEDLINE=21874127; PubMed=11867749;
RA Shi Y., Hoi M., Evans R.M.;
RT "The peroxisome proliferator-activated receptor delta, an integrator
of transcriptional repression and nuclear receptor signaling.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618 (2002).
RN [7]
RN FUNCTION, AND INTERACTION WITH REPSUH.
RP MEDLINE=22261914; PubMed=12374742;
RA Oswald F., Kostezka U., Astrahantseff K., Bourteele S., Dillinger K.,
RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Lipray S.,
RA Schmid R.M.;

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U67593; AAB99541.1; -.
 DR PIR; A64490; A64490.
 DR TIGR; MJ1522; -.

DR InterPro; IPR000878; Cor/por Metransf.
 DR Pfam; PF005090; TP methylase; 1.
 KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
 KW Methyltransferase; Complete proteome.
 SQ SEQUENCE 211 AA; 23805 MW; 279A1A2B14369510 CRC64;

Query Match 40.6%; Score 43; DB 1; Length 211;
 Best Local Similarity 54.5%; Pred. No. 23;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 7 VVGLSPGGEQY 17
 :||: ||:|
 Db 4 IVGIGGDEY 14

RESULT 9

ID MER METTI STANDARD; PRT; 326 AA.
 AC Q9UXP0;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Coenzyme F420-dependent N(5),N(10)-methylentetrahydromethanopterin
 DE reductase (EC 1.5.99.11) (Methylene-H(4)MPT reductase).
 GN MER OR FFDA.
 OS Methanobolus tindarius.
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
 OC Methanosarcinaceae; Methanobolus.
 OX NCBI_TaxID=2221;
 RN [1]

SEQUENCE FROM N.A.

RC STRAIN=DSM 2278;
 RC MEDLINE=99132696; PubMed=9933933;
 RA Wostenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
 RA Gottschalk G., Blaut M.;
 RA "The F420H2-dehydrogenase from Methanobolus tindarius: cloning of the
 RT ffd operon and expression of the genes in *Escherichia coli*.";
 RL FEMS Microbiol. Lett. 170:389-398(1999).
 CC -!- FUNCTION: Catalyzes the reversible reduction of methylene-H(4)MPT
 CC to methyl-H(4)MPT (By similarity).
 CC -!- CATALYTIC ACTIVITY: N(5),N(10)-methylentetrahydromethanopterin +
 CC reduced coenzyme F420 = 5-methyl-5,6,7,8-tetrahydromethanopterin +
 CC coenzyme F420.
 CC -!- PATHWAY: Methanogenesis from carbon dioxide; fifth step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the mer family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AJ011519; CAB56639.1; -.
 DR PIR; T45226; T45226.
 DR HAMAP; MF_01091; -.
 DR InterPro; IPR002103; Bac luciferase.
 DR Pfam; PF00296; Bac luciferase; 1.
 KW Methanogenesis; One-carbon metabolism; Oxidoreductase.
 SQ SEQUENCE 326 AA; 34043 MW; 16F3AB9733A45D82 CRC64;

Query Match 40.6%; Score 43; DB 1; Length 326;
 Best Local Similarity 53.3%; Pred. No. 35;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAGLPVVVGLSPGGEQ 15
 :||: ||:|
 Db 83 ISGGRALUGLPGGEQ 97

RESULT 10

ID COBL MYCTU STANDARD; PRT; 390 AA.
 AC Q10671;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Precorrin-6Y C5,15-methyltransferase [decarboxylating] (BC 2.1.1.132)
 DE (Precorrin-6 methyltransferase) [Precorrin-6Y methylase].
 GN COBL OR RV2072C OR MT2132 OR MTCV49.11C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]

SEQUENCE FROM N.A.

RC STRAIN=H37Rv;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultson J.E., Taylor K., Whitehead S., Barrrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]

SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;
 RC MEDLINE=22206494; PubMed=12218036;
 RA Fleischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickley B.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Decher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 CC -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
 CC PRECORRIN-6Y TO FORM PRECORRIN-8X.
 CC -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
 CC S-adenosyl-L-homocysteine + precorrin-8X + CO(2).
 CC -!- PATHWAY: Cobalamin biosynthesis.
 CC -!- SIMILARITY: TO S-TYPHIMURIUM CBIE; ALSO, LOW, TO OTHER
 CC METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; Z73966; CAA98225.1; -.
 DR EMBL; AE007063; AAK46412.1; -.
 DR PIR; C70765; C70765.
 DR TIGR; MT2132; -.
 DR InterPro; IPR006365; Cobl.
 DR InterPro; IPR000978; Cor/por Metransf.


```

DR InterPro: IPR000051; SAM bind.
DR Pfam: PF00590; TP_methylase; 1.
DR TIGRFAMs: TIGR01468; coBL_cblE; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Complete proteome.
FT CONFLICT 205 205 L -> P (IN REF. 2).
FT CONFLICT 327 327 L -> H (IN REF. 2).
SQ SEQUENCE 390 AA; 41854 MW; FB42EPF7562F21F3 CRC64;

Query Match 40.6%; Score 43; DB 1; Length 390;
Best Local Similarity 60.0%; Pred. No. 43;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 LPNAVGLSPGQVEYH 18
DB 55 LPNAVGLSPDGADLH 69

RESULT 11
YTSJ_CABEL STANDARD; PRT; 505 AA.
AC P90838;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0027 protein F16A11.2 in chromosome I.
GN F16A11.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Durbini R.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RA Durbini R.;
RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the UPF0027 (rtCB) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z81505; CAB04121.2; -.
CC WormPep; F16A11.2; CE23663.
CC InterPro: IPR001233; UPF0027.
CC Pfam: PF01139; UPF0027; 1.
CC PROSITE: PS01288; UPF0027; 1.
KW Hypothetical protein.
SQ SEQUENCE 505 AA; 55230 MW; D528F702E2596909 CRC64;

Query Match 40.6%; Score 43; DB 1; Length 505;
Best Local Similarity 45.0%; Pred. No. 56;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 MAGLPNAVGLSPGQVEYHRG 20
DB 72 VASLPGVGHSLGLFDIHS 91

RESULT 12
C4AD_DROME STANDARD; PRT; 516 AA.
ID C4AD_DROME
AC Q9V4T3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable cytochrome P450 4ad1 (EC 1.14.-.-) (CYP1VAD1).
CN CYP4ADI OR CG2110.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chertny J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RC MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan X.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -!- FUNCTION: May be involved in the metabolism of insect hormones and
CC in the breakdown of synthetic insecticides (By similarity).
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
CC (Potential).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE003837; AAF59092.1; -.
CC EMBL; AY061058; AAL28606.1; -.
CC HSSP; F14779; IJFZ.
CC FlyBase; FBgn0033292; Cyp4ad1.
CC Pfam; PF00067; p450; 1.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum; Hypothetical protein.
FT METAL 445 445 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 516 AA; 58970 MW; 648EA22492AF58C7 CRC64;

Query Match 40.6%; Score 43; DB 1; Length 516;
Best Local Similarity 56.2%; Pred. No. 57;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 LPVVVGLSPGGEQYHR 19
DB 469 LPVVVGLSPGGINDSR 484

RESULT 13
CH60 PROAC
ID CH60 PROAC STANDARD; PRT; 544 AA.
AC Q9K2U4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein) (Heat shock protein
DE 60).
GN GROEL OR GROEL OR HSP60.
OS Propionibacterium acnes.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Propionibacterineae; Propionibacteriaceae; Propionibacterium.
OX NCBI_TaxID=1747;
RN [1]
RP SEQUENCE FROM N.A.
RA Cho Y.;
RT "Propionibacterium acnes hsp60 gene.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=P37;
RA Farrar M.D., Ingham E., Holland K.T.;
RT "Cloning and sequencing of a groEL homolog from Propionibacterium
RT acnes.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -!- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB036414; BAA92382.1; -.
CC EMBL; AF220611; AAF33788.1; -.
CC HSPSP; P06139; IGRN.
CC HAMAP; MF_00600; -. 1.
CC InterPro; IPR001844; Chaperonin Cpn60.
CC InterPro; IPR002423; Cpn60/Tcf-1.
CC InterPro; IPR008950; GroEL-ATPase.

or send an email to license@isb-sib.ch).
-----
Pfam; PF00119; cpn60_TCF1; 1.
PRINTS; PR00298; CHAPERONIN60.
DR PROSITE; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding
SQ SEQUENCE 544 AA; 56839 MW; 9B4314A8BF8A94DC CRC64;

Query Match 40.6%; Score 43; DB 1; Length 544;
Best Local Similarity 52.9%; Pred. No. 60;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGLPVVVGLSPGGEQY 17
DB 466 VAGLPAGQGLNAANDEY 482

RESULT 14
HNFA MOUSE
ID HNFA MOUSE STANDARD; PRT; 628 AA.
AC P22361;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hepatocyte nuclear factor 1-alpha (HNF-1A) (Liver-specific
DE transcription factor 1F-B1) (LFB1).
GN TCF1 OR HNF1A OR HNF-1A OR HNF-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9108607; PubMed=2263635;
RA Kuo C.J., Conley P.B., Hsieh C.L., Francke U., Crabtree G.R.;
RT "Molecular cloning, functional expression, and chromosomal
RT localization of mouse hepatocyte nuclear factor 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9838-9842(1990).
CC -!- FUNCTION: Required for the expression of several liver specific
CC genes. Binds to the inverted palindrome 5'-GTTAATNATTAAAC-3'.
CC -!- SUBUNIT: Binds DNA as a dimer.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Liver.
CC -!- SIMILARITY: Belongs to the HNF1 homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M57966; AAA37821.1; -.
CC PIR; A39262; A39262.
CC PDB; 1F93; 20-SEP-00.
CC PDB; 1G2Y; 17-JAN-01.
CC PDB; 1G2Z; 17-JAN-01.
CC PDB; 1G39; 17-JAN-01.
CC PDB; 1JB6; 11-JUL-01.
CC PDB; 1LFB; 31-OCT-93.
CC TRANSFAC; T01211; -.
CC MGD; MGI:98504; Cncl.
CC GO; GO:0005634; C:nucleus; IDA.
CC InterPro; IPR006899; HNF-1_N.
CC InterPro; IPR006898; HNF1_C.
CC InterPro; IPR001356; HNF1b_C.
CC InterPro; IPR006897; Homeobox.
CC Pfam; PF04814; HNF-1_N; 1.
CC Pfam; PF04813; HNF-1A_C; 1.
CC Pfam; PF04812; HNF-1B_C; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
```

DR PROSITE; PS50071; HOMEBOX 2; 1.
 KW Transcription regulation; DNA-binding; Homeobox; Nuclear protein;
 Activator; Trans-acting factor; 3D-structure.
 FT DOMAIN 1 31 DIMERIZATION.
 FT DOMAIN 71 80 ASP/GLU-RICH (ACIDIC) (POTENTIAL
 INVOLVEMENT WITH TRANSCRIPTION).
 FT DOMAIN 197 205 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 199 279 HOMEBOX.
 FT DNA BIND 238 258 21 AMINO ACID LOOP BETWEEN HELIX 2 AND 3.
 SQ SEQUENCE 628 AA; 67237 MW; 737920D1A369B9DD CRC64;

Query Match 40.6%; Score 43; DB 1; Length 628;
 Best Local Similarity 57.1%; Pred. No. 70;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAGLEFAVVGSLSGE 14
 |||||:|:|
 404 MASLPQVMTIGDGE 417

Db

RESULT 15
 HNFA RAT STANDARD; PRT; 628 AA.
 ID P15257;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hepatocyte nuclear factor 1-alpha (HNF-1A) (liver-specific
 transcription factor LF-B1) (LFBI).
 DE TCF1 OR HNF1A OR HNF-1A OR HNF-1.
 OS Rattus norvegicus (Rat).
 GN Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=90003224; PubMed=2571419;
 RA Frain M., Swart G., Monaci P., Nicosia A., Staempfli S., Frank R.,
 Cortese R.;
 RT "The liver-specific transcription factor LF-B1 contains a highly
 diverged homeobox DNA binding domain.";
 RL Cell 59:145-157(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91016926; PubMed=2216777;
 RA Chouard T., Blumenfeld M., Bach I., Vandekerckhove J., Cereghini S.,
 Yaniv M.;
 RT "A distal dimerization domain is essential for DNA-binding by the
 atypical HNF1 homeobox domain.";
 RL Nucleic Acids Res. 18:5853-5863(1990).
 RN [3]
 RP SEQUENCE OF 166-628 FROM N.A.
 RX MEDLINE=90249741; PubMed=1970973;
 RA Baumhueter S., Mendel D.B., Conley P.B., Kuo C.J., Turk C.,
 Graves M.K., Edwards C.A., Courtois G., Crabtree G.R.;
 RT "HNF-1 shares three sequence motifs with the POU domain proteins and
 is identical to LF-B1 and APF.";
 RL Genes Dev. 4:372-379(1990).
 RN [4]
 RP SEQUENCE OF 1-12 FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RA Tomei L., Piaggio G., Toniatti C., Lazzaro D., de Francesco R.,
 Pozzi L., Gerstner J., Cortese R.;
 RL Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP POSITION OF HOMEBOX.
 RX MEDLINE=90106643; PubMed=1967225;
 RA Finney M.;
 RT "The homeodomain of the transcription factor LF-B1 has a 21 amino
 acid loop between helix 2 and helix 3.";
 RL Cell 60:5-6(1990).
 RN [6]

RP STRUCTURE BY NMR OF 1-32.
 RX MEDLINE=91105074; PubMed=1988016;
 RA Pastore A., de Francesco R., Barbato G., Castiglione Morelli M.A.,
 Motta A., Cortese R.;
 RT "1H resonance assignment and secondary structure determination of the
 dimerization domain of transcription factor LFBI.";
 RL Biochemistry 30:148-153(1991).
 RN [7]
 RP STRUCTURE BY NMR OF 195-286.
 RX MEDLINE=93259120; PubMed=8491172;
 RA Leitig B., de Francesco R., Tomei L., Cortese R., Otting G.,
 Wuehrich K.;
 RT "The three-dimensional NMR-solution structure of the polypeptide
 fragment 195-286 of the LFBI/HNF1 transcription factor from rat liver
 comprises a nonclassical homeodomain.";
 RL EMBO J. 12:1797-1803(1993).
 RN [8]
 RP STRUCTURE BY NMR OF 195-286.
 RX MEDLINE=97272000; PubMed=9126845;
 RA Schott O., Billeter M., Leitig B., Wider G., Wuehrich K.;
 RT "The NMR solution structure of the non-classical homeodomain from the
 rat liver LFBI/HNF1 transcription factor.";
 RL J. Mol. Biol. 267:673-683(1997).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 195-286.
 RX MEDLINE=93259121; PubMed=8491173;
 RA Ceska T.A., Tamers M., Monaci P., Nicosia A., Cortese R., Suck D.;
 RT "The X-ray structure of an atypical homeodomain present in the rat
 liver transcription factor LFBI/HNF1 and implications for DNA
 binding.";
 RL EMBO J. 12:1805-1810(1993).
 CC -!- FUNCTION: Required for the expression of several liver specific
 genes. Binds to the inverted palindrome 5'-GTTAATNATTAAAC-3'.
 CC -!- SUBUNIT: Binds DNA as a dimer.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Liver.
 CC -!- SIMILARITY: Belongs to the HNF1 homeobox family.
 CC -!- SIMILARITY: Contains 1 homeobox domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J03170; AAA41524.1; -;
 DR EMBL; X54423; CAA38295.1; -;
 DR EMBL; X67649; CAA47891.1; -;
 DR EMBL; X53297; CAA37387.1; ALT_INIT.
 DR PIR; A33333; A33333.
 DR PIR; S25485; S25485.
 DR PDB; 1LFB; 31-OCT-93.
 DR PDB; 2LFB; 12-MAR-97.
 DR TRANSFAC; T00369; -;
 DR InterPro; IPR006899; HNF-1_N.
 DR InterPro; IPR006898; HNF1A_C.
 DR InterPro; IPR006897; HNF1B_C.
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF04814; HNF-1_N; 1.
 DR Pfam; PF04813; HNF-1A_C; 1.
 DR Pfam; PF04812; HNF-1B_C; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PS50071; HOMEBOX_2; 1.
 DR Transcription regulation; DNA-binding; Homeobox; Nuclear protein;
 KW Activator; Trans-acting factor; 3D-structure.
 FT DOMAIN 1 31 DIMERIZATION.
 FT DOMAIN 71 80 ASP/GLU-RICH (ACIDIC) (POTENTIAL
 INVOLVEMENT WITH TRANSCRIPTION).
 FT DOMAIN 197 205 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 199 279 HOMEBOX.

FT DOMAIN 238 258 21 AMINO ACID LOOP BETWEEN HELIX 2 AND 3.
 FT HELIX 208 218
 FT TURN 219 220
 FT HELIX 226 241
 FT TURN 242 245
 FT TURN 248 249
 FT TURN 251 254
 FT HELIX 255 257
 FT HELIX 261 273
 FT TURN 274 275
 SQ SEQUENCE 628 AA; 67213 MW; 8D28099308C86A52 CRC64;
 Query Match 40.6%; Score 43; DB 1; Length 628;
 Best Local Similarity 57.1%; Pred. No. 70;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MAGLPVVGLSPGE 14
 Db ||||| : |||
 404 MASLPQVMTIGFGE 417

Search completed: May 7, 2004, 12:34:31
 Job time : 8.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 37.4 Seconds
(without alignments)

168.726 Million cell updates/sec

Title: US-09-786-214A-9

Perfect score: 106

Sequence: 1 MAGLPAVVGLSPGEQVHRG 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTEMPL 25:*

2: SP_arChaea:*

3: SP_bacteria:*

4: SP_fungi:*

5: SP_invertebrate:*

6: SP_mammal:*

7: SP_mhc:*

8: SP_organelle:*

9: SP_phase:*

10: SP_plant:*

11: SP_rodent:*

12: SP_virus:*

13: SP_vertibrate:*

14: SP_unclassified:*

15: SP_rvirus:*

16: SP_bacteriap:*

17: SP_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	60.4	41	11	Q8k408 rattus norv
2	52	49.1	150	16	Q8PE55 xanthomonas
3	50	47.2	242	16	Q89Y25 bradyrhizob
4	50	47.2	277	16	Q9KQK6 vibrio chol
5	50	47.2	284	16	Q7VF15 helicobacte
6	50	47.2	381	16	Q9RRL5 deinococcus
7	49	46.2	277	16	Q87RD3 vibrio para
8	49	46.2	4823	13	Q93321 fugu rubrip
9	48	45.3	192	15	Q8UTQ8 human immu
10	48	45.3	192	15	Q90D29 human immu
11	48	45.3	278	16	Q9CMF6 pasteurella
12	48	45.3	300	16	Q8EMZ7 oceanobacil
13	48	45.3	485	10	Q7XNE4 oryza sativ
14	47	44.3	192	15	Q994Q2 human immu
15	47	44.3	192	15	Q9QSR2 human immu
16	47	44.3	192	15	Q994B2 human immu

17	47	44.3	250	16	Q8D0G6
18	47	44.3	280	16	Q8ZP42
19	47	44.3	280	16	Q8XDC4
20	47	44.3	280	16	Q8Z7F3
21	47	44.3	289	16	Q8XBG2
22	47	44.3	293	16	O25975
23	47	44.3	293	16	Q9ZJ12
24	47	44.3	390	16	Q98CK1
25	47	44.3	661	10	Q8S8J8
26	46.5	43.9	242	16	Q7WCX1
27	46.5	43.9	298	16	Q7W5D6
28	46.5	43.9	298	16	Q7VTI2
29	46.5	43.9	563	16	Q9X8S9
30	46	43.4	140	16	Q8P5G3
31	46	43.4	143	4	Q8NAW2
32	46	43.4	192	15	Q90CX5
33	46	43.4	192	15	Q8UTJ5
34	46	43.4	192	15	O70887
35	46	43.4	274	16	Q9PPC9
36	46	43.4	300	16	O34990
37	46	43.4	462	16	Q9KZB0
38	46	43.4	753	16	Q89T31
39	45	42.5	85	6	O77786
40	45	42.5	192	15	Q90ML0
41	45	42.5	192	15	Q994N4
42	45	42.5	192	15	Q91W51
43	45	42.5	192	15	Q9WQH9
44	45	42.5	192	15	Q9WP98
45	45	42.5	192	15	Q99BN6

ALIGNMENTS

RESULT 1

Q8K408 PRELIMINARY; PRT; 41 AA.

AC Q8K408; 22, Created)

DT 01-OCT-2002 (TREMBLrel. 22, last sequence update)

DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)

DE Truncated macrophage colony stimulating factor.

GN CSF1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LEW.tl;

RX MEDLINE=22069908; PubMed=12074592;

RA Dobbins D.E., Sood R., Hashimoto A., Hansen C.T., Wilder R.L.,

RA Remmers E.F.;

RT "Mutation of macrophage colony stimulating factor (CSF1) causes

osteopetrosis in the tl rat."

RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).

DR EMBL; AF514357; AAM54137.1; -

SQ SEQUENCE 41 AA; 4178 MW; 1D342C19BD18AA41 CRC64;

Query Match 60.4%; Score 64; DB 11; Length 41;

Best Local Similarity 70.0%; Pred. No. 0.014;

Matches 14; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MAGLPAVVGLSPGEQVHRG 20

DB 18 MDGLPAAAGLSPREQCCRG 37

RESULT 2

Q8PE55 PRELIMINARY; PRT; 150 AA.

ID Q8PE55

AC Q8PE55; 22, Created)

DT 01-OCT-2002 (TREMBLrel. 22, last sequence update)

DR EMBL; AP005935; BAC45395.1; -.
DR InterPro: IPR005493; Methyltransf 6.

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Formyltetrahydrofolate deformylase PurU (BC 3.5.1.10).
GN PUKU OR HH1691.
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Stenzenbach T., Drescher B., Brandt P.,
RA Bell M., Droge M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
RW EMBL; AE017149; AAP78288.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 284 AA; 32660 MW; AAED2DC5086236E3 CRC64;

Query Match 47.2%; Score 50; DB 16; Length 284;
Best Local Similarity 52.9%; Pred. No. 17;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 LPVVGLSPGQEQYHRG 20
DB 197 LPAFIGANPYQAYERG 213

RESULT 6
Q9RRL5 ID Q9RRL5 PRELIMINARY; PRT; 381 AA.
AC Q9RRL5
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450.
DN DR2473.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Ueberback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Manton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AE002076; AAF12016.1; -.
DR PIR; F75270; F75270.
DR TIGR; DR2473; -.
DR GO; GO:0004497; F:monoxygenase activity; IEA.
DR GO; GO:0006118; F:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME P450; 1.
KW Heme; Monoxygenase; Oxidoreductase; Complete proteome.
SQ SEQUENCE 381 AA; 41940 MW; F191EA69F1797B53 CRC64;

Query Match 47.2%; Score 50; DB 16; Length 381;

Best Local Similarity 100.0%; Pred. No. 23;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GLPAVVGSLP 12
DB 51 GLPAVVGSLP 60

RESULT 7
Q87RD3 ID Q87RD3 PRELIMINARY; PRT; 277 AA.
AC Q87RD3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Formyltetrahydrofolate deformylase.
DN VP0864.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae."
RL Lancet 361:743-749(2003).
RW EMBL; AP005075; BAC59127.1; -.
DR GO; GO:0016597; F:amino acid binding; IEA.
DR GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
DR GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. .; IEA.
DR GO; GO:0006189; P:'de novo' IMP biosynthesis; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR002376; formyl_transf.
DR InterPro; IPR004810; PurU.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00551; formyl_transf; 1.
DR PRINTS; PR01575; FFH4HYDLASE.
KW Complete proteome.
SQ SEQUENCE 277 AA; 31620 MW; D5F0712E1537CF4C CRC64;

Query Match 46.2%; Score 49; DB 16; Length 277;
Best Local Similarity 52.9%; Pred. No. 23;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 LPVVGLSPGQEQYHRG 20
DB 190 LPAFIGAKPYQAYDRG 206

RESULT 8
Q93321 ID Q93321 PRELIMINARY; PRT; 4823 AA.
AC Q93321
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE All-1 related protein.
DN ALR.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;

```

RA Gellner K., Brenner S.;
 RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
 rubripes."; 9:251-258(1999).
 RL Genome Res. 9:251-258(1999).
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 DR EMBL; AF056116; AAC34383.1; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR003889; FYrich_C.
 DR InterPro; IPR003888; FYrich_N.
 DR InterPro; IPR000910; HM12_box.
 DR InterPro; IPR003616; PostSET.
 DR InterPro; IPR001214; SET.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00628; PHD; 5.
 DR Pfam; PF00856; SET; 1.
 DR SMART; SM00542; FYRC; 1.
 DR SMART; SM00541; FYRN; 1.
 DR SMART; SM00398; HMG; 1.
 DR SMART; SM00249; PHD; 8.
 DR SMART; SM00508; PostSET; 1.
 DR SMART; SM00184; RING; 4.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS0868; POSTSET; 1.
 DR PROSITE; PS0280; SET; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS00016; ZF_PHD_2; 1.
 DR PROSITE; PS00089; ZF_RING_2; 1.
 DR PROSITE; PS00089; ZF_RING_2; 1.
 SQ SEQUENCE 4823 AA; 526260 MW; BD0C5F4EAD0F9C07 CRC64;

Query Match 46.2%; Score 49; DB 13; Length 4823;
 Best Local Similarity 66.7%; Pred. No. 5.2e+02;
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 PAVVGLSPGGEQYHR 19
 DB 1900 PALGGLSPSELEKX 1914
 |||:|||||:|

RESULT 9
 Q8UTQ8 PRELIMINARY; PRT; 192 AA.
 ID Q8UTQ8
 AC Q8UTQ8;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Vif protein (Vifon infectivity factor) (SOR protein).
 GN VIF.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=98BWC14.a3;
 RA Novitsky V.A., Smith U.R., Gilbert P., McLane M.P., Chigwedere P.,
 RA Williamson C., Ndung'u T., Klein I., Chang S.-Y., Peter T., Thior I.,
 RA Foley B.T., Gaolekwe S., Rybak N., Gaseitsiwe S., Vannberg F.,
 RA Marlink R., Lee T.-H., Essex M.;
 RA "HIV-1 subtype C molecular phylogeny: consensus sequence for an AIDS
 RT vaccine design."
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RL -1- FUNCTION: DETERMINES VIRUS INFECTIVITY (BY SIMILARITY).
 CC EMBL; AF443078; AAL34591.1; -.
 DR GO; GO:0019058; P:viral infectious cycle; IEA.
 DR InterPro; IPR00475; Viral_infect.
 DR Pfam; PF00559; Vif; 1.
 DR PRINTS; PR00349; VIRIONINFECT.
 DR ProDom; PD000063; Viral_infect; 1.
 DR HSP; P08179; 1GRC.
 DR GO; GO:0016597; F:amino acid binding; IEA.

Query Match 45.3%; Score 48; DB 15; Length 192;
 Best Local Similarity 66.7%; Pred. No. 22;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 9 GLSPGGEQYHRG 20
 DB 71 GLQPGGEREWHLG 82
 |||:|||||:|

RESULT 10
 Q90D29 PRELIMINARY; PRT; 192 AA.
 ID Q90D29
 AC Q90D29;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Vif protein (Vifon infectivity factor) (SOR protein).
 GN VIF.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=97TZ01;
 RX MEDLINE=21395692; PubMed=11504977;
 RA Hoelscher M., Kim B., Maboko L., Mhalu F., von Sonnenburg F.,
 RA Birx D.L., McCutchan F.E.;
 RA the UNAIDS Network for HIV Isolation Characterization.;
 RT "High proportion of unrelated HIV-1 intersubtype recombinants in the
 RT Mbeya region of southwest Tanzania."
 RL AIDS 15;1461-1470(2001).
 CC -1- FUNCTION: DETERMINES VIRUS INFECTIVITY (BY SIMILARITY).
 DR EMBL; AF361871; AAK94212.1; -.
 DR GO; GO:0019058; P:viral infectious cycle; IEA.
 DR InterPro; IPR000475; Viral_infect.
 DR Pfam; PF00559; Vif; 1.
 DR PRINTS; PR00349; VIRIONINFECT.
 DR ProDom; PD000063; Viral_infect; 1.
 DR AIDS.
 SQ SEQUENCE 192 AA; 22674 MW; 11B799C5DEA99F77 CRC64;

Query Match 45.3%; Score 48; DB 15; Length 192;
 Best Local Similarity 66.7%; Pred. No. 22;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 9 GLSPGGEQYHRG 20
 DB 71 GLQPGGEREWHLG 82
 |||:|||||:|

RESULT 11
 Q9CMF6 PRELIMINARY; PRT; 278 AA.
 ID Q9CMF6
 AC Q9CMF6;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PurU.
 GN PURU OR PM0873.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RA "Complete genomic sequence of Pasteurella multocida Pm70."
 RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 RL EMBL; AE006126; AAK02957.1; -.
 DR HSP; P08179; 1GRC.
 DR GO; GO:0016597; F:amino acid binding; IEA.

```

DR GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
DR GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. . .; IEA.
DR GO; GO:0006189; P:'de novo' IMP biosynthesis; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR002376; formyl_transf.
DR InterPro; IPR004810; PurU.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00551; formyl_transf; 1.
DR PRINTS; PR01575; FFH4HYDRASE.
DR TIGRfams; TIGR00655; PurU; 1.
KW Complete proteome.
SQ SEQUENCE 278 AA; 32086 MW; F303E499DFFA0B70 CRC64;

Query Match 45.3%; Score 48; DB 16; Length 278;
Best Local Similarity 52.9%; Pred. No. 33;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 LPVVGLSPGQEQYHRG 20
Db 191 LPAFIGAKPYHQAYERG 207

RESULT 12
Q8EMZ7 PRELIMINARY; PRT; 300 AA.
AC Q8EMZ7
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Formyltetrahydrofolate deformylase (EC 3.5.1.10).
GN OB2693.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AF004602; BAC14649.1;
DR GO; GO:0016597; F:amino acid binding; IEA.
DR GO; GO:0008864; F:formyltetrahydrofolate deformylase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0016742; F:hydroxymethyl-, formyl- and related transfe. . .; IEA.
DR GO; GO:0006189; P:'de novo' IMP biosynthesis; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR002376; formyl_transf.
DR InterPro; IPR004810; PurU.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00551; formyl_transf; 1.
DR PRINTS; PR01575; FFH4HYDRASE.
DR TIGRfams; TIGR00655; PurU; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 300 AA; 34894 MW; D44D0FECC596BF2 CRC64;

Query Match 45.3%; Score 48; DB 16; Length 300;
Best Local Similarity 52.9%; Pred. No. 36;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 LPVVGLSPGQEQYHRG 20
Db 212 LPAFIGAKPYERAYDRG 228

RESULT 13
Q7XNE4 PRELIMINARY; PRT; 485 AA.
ID Q7XNE4
```

```

AC Q7XNE4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSUNBA0009P12.18 protein.
GN OSUNBA0009P12.18.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Yang K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.O., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Hu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL662952; CAE04133.1;
SQ SEQUENCE 485 AA; 52334 MW; F7685952D5538456 CRC64;

Query Match 45.3%; Score 48; DB 10; Length 485;
Best Local Similarity 60.0%; Pred. No. 61;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 LPVVGLSPGQEQYH 18
Db 317 LPALSKLSPGQAAHY 331

RESULT 14
Q994Q2 PRELIMINARY; PRT; 192 AA.
ID Q994Q2
AC Q994Q2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 24, Last annotation update)
DE Vif protein (Viron infectivity factor) (SOR protein).
GN VIF.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96ZM753;
RX MEDLINE=21094715; PubMed=11177395;
RA Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA Kalish M.L., Shaw G.M., Allen S., Hahn B.H., Gao F.;
RT "Near full-length clones and reference sequences for subtype C
RT isolates for HIV type 1 from three different continents."
RL AIDS Res. Hum. Retroviruses 17:161-168(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=96ZM751;
RA Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA Allen S., Shaw G.M., Hahn B.H., Gao F.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: DETERMINES VIRUS INFECTIVITY (BY SIMILARITY).
DR EMBL; AF286225; AAK30974.1;
DR GO; GO:0019058; P:Viral infectious cycle; IEA.
DR InterPro; IPR000475; Viral_infect.
DR Pfam; PF00559; Vif; 1.
DR PRINTS; PR00349; VIRIONINFECT.
DR Prodrom; PD000063; Viral_infect; 1.
KW AIDS.
SQ SEQUENCE 192 AA; 22727 MW; 55E01D4BBBCD93DC6 CRC64;

Query Match 44.3%; Score 47; DB 15; Length 192;
```


Best Local Similarity 66.7%; Pred. No. 32;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 9 GLSPGGEYHRG 20
|||:|:|
Db 71 GLHFGEREHLG 82

RESULT 15

Q9QSR2 PRELIMINARY; PRT; 192 AA.
AC Q9QSR2;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Vif protein (Viron infectivity factor) (SOR protein).
GN VIF.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V1850;
RX MEDLINE=20192166; PubMed=10725202;
RA Laukkanen T., Carr J.K., Janssens W., Liitsola K., Gotte D.,
RA McCutchan F.E., Op de Coul E., Cornelissen M., Heyndrickx L.,
RA van der Groen G., Salminen M.O.;
RT "Virtually full-length subtype F and F/D recombinant HIV-1 from Africa
and South America.";
RL Virology 269:95-104(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=V1850;
RA Carr J.K., Kim B., Sanders-Buell E., Salminen M.O., Alaeus A.,
RA Albert J.A., Birx D.B., McCutchan F.E.;
RT "HIV-1 isolate V1850 from Zaire, complete genome."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DETERMINES VIRUS INFECTIVITY (BY SIMILARITY).
DR EMBL; AF077336; AAD46089.1; -;
DR GO; GO:0019058; P:Viral infectious cycle; IEA.
DR InterPro; IPR000475; Viral_infect.
DR Pfam; PF00559; Vif; 1.
DR PRINTS; PR00349; VIRIONINFECT.
DR ProDom; PD000063; Viral_infect; 1.
KW AIDS.
SQ SEQUENCE 192 AA; 22629 MW; F165805BFCD4427 CRC64;

Query Match 44.3%; Score 47; DB 15; Length 192;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 9 GLSPGGEYHRG 20
|||:|:|
Db 71 GLHFGEREHLG 82

Search completed: May 7, 2004, 12:37:52
Job time : 39.5667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 55.8 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-9

Perfect score: 106

Sequence: 1 MAGLPAAVVLGSGEGEYHRG 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Geneseqp29Jan04:*
- 2: Geneseqp1980s:*
- 3: Geneseqp1990s:*
- 4: Geneseqp2000s:*
- 5: Geneseqp2001s:*
- 6: Geneseqp2002s:*
- 7: Geneseqp2003as:*
- 8: Geneseqp2003bs:*
- 9: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	20	3	AAy84265 Truncated
2	106	100.0	25	3	AAy84264 Peptide o
3	75	70.8	15	3	AAy84269 Peptide d
4	72	67.9	14	3	AAy84266 Peptide d
5	68	64.2	13	3	AAy84267 Peptide d
6	65	61.3	13	3	AAy84268 Peptide d
7	50	47.2	277	6	ABu49441 Protein e
8	49	46.2	191	3	AAy84265 HIV-1 non
9	49	46.2	278	6	ABu30626 Protein e
10	48	45.3	278	6	ABU39195 Protein e
11	48	45.3	287	6	ABu17217 Protein e
12	48	45.3	303	6	ADa36829 Acinetoba
13	47	44.3	151	5	ABu50765 Helicobac
14	47	44.3	155	5	ABu51750 Helicobac
15	47	44.3	192	3	AAy84269 HIV-1 non
16	47	44.3	238	6	ABu27483 Protein e
17	47	44.3	278	6	ABu44771 Protein e
18	47	44.3	280	6	ABu48054 Protein e
19	47	44.3	280	6	ABU31449 Protein e
20	47	44.3	280	6	ABU15054 Protein e
21	47	44.3	280	6	ABU47098 Protein e
22	47	44.3	282	6	ABU41187 Protein e
23	47	44.3	293	6	ABM69037 Photorhab
24	47	44.3	293	2	AAW98485 H. pylori
25	47	44.3	293	6	ABU31073 Protein e

26	46	43.4	143	5	AAO21677	AAO21677 Human sec
27	46	43.4	143	7	ADB64633	ADB64633 Human pro
28	46	43.4	187	6	ABU21296	ABu21296 Protein e
29	46	43.4	192	3	AAy84269	AAy84269 HIV-1 non
30	46	43.4	234	4	AAy84268	AAy84268 Human imm
31	46	43.4	274	6	ABU26498	ABu26498 Protein e
32	46	43.4	300	2	AAy16108	AAy16108 A formate
33	46	43.4	300	2	AAy16108	AAy16108 A formate
34	45	42.5	116	4	ABG24038	ABg24038 Novel hum
35	45	42.5	287	7	ADB80057	ADB80057 Mycobacte
36	45	42.5	334	4	ABU11182	ABu11182 Human gua
37	45	42.5	334	4	AAy79690	AAy79690 Human pro
38	45	42.5	403	4	AAy40847	AAy40847 Human pol
39	45	42.5	456	4	AAy78706	AAy78706 Human pro
40	44	41.5	506	4	AAy39061	AAy39061 Human pol
41	44	41.5	565	4	ABY71810	ABy71810 Drosophila
42	44	41.5	579	4	AAy76817	AAy76817 Corynebact
43	44	41.5	940	4	AAy90917	AAy90917 C glutami
44	44	41.5	968	4	AAU28194	AAu28194 Novel hum
45	44	41.5	1663	6	ABO53095	ABO53095 Novel hum

ALIGNMENTS

RESULT 1

AAy84265	ID	AAy84265 standard; peptide; 20 AA.
XX	AC	AAy84265;
XX	DT	12-JUL-2000 (first entry)
XX	DE	Truncated macrophage colony stimulating factor tumour antigen.
XX	DE	tumour rejection antigen; macrophage colony stimulating gene;
KW	KW	macrophage-colony stimulating factor; antigen presenting cell;
KW	KW	human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX	OS	Homo sapiens.
XX	PN	WO200013699-A1.
XX	PD	16-MAR-2000.
XX	PF	03-SEP-1999; 99WO-US020344.
XX	PR	04-SEP-1998; 98US-0099077P.
XX	PA	(LUDW-) LUDWIG INST CANCER RES.
PI	PI	Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
DR	DR	WPI; 2000-256859/22.
DR	DR	N-PSDB; AA299675.
XX	XX	Isolated polypeptide used to treat subjects having a disorder
PT	PT	characterized by expression of alternative open reading frame macrophage-
PT	PT	colony stimulating factor comprises 25 amino acid residue sequence.
XX	XX	Claim 3; Page 64; 74pp; English.
PS	PS	The present sequence represents a truncated tumour rejection antigen
XX	XX	precursor, and is encoded by a truncated alternative open reading frame
CC	CC	(ORF) of human macrophage colony stimulating gene. Peptides derived from
CC	CC	the alternative ORF of macrophage-colony stimulating factor, when
CC	CC	presented by an antigen presenting cell having a human leukocyte antigen
CC	CC	(HLA) class I molecule, effectively induce the activation and
CC	CC	proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC	CC	acids derived from the alternate ORF of macrophage-colony stimulating
CC	CC	factor are useful for enriching selectively a population of T lymphocytes
CC	CC	with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC	CC	characterized by expression of the polypeptide, and for identifying

CC functional variants and mimetics

```

XX Sequence 20 AA;
SQ
  Query Match      100.0%; Score 106; DB 3; Length 20;
  Best Local Similarity 100.0%; Pred. No. 8.7e-09;
  Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGLPAVVGSLSPGQEQYHRG 20
Db 1 MAGLPAVVGSLSPGQEQYHRG 20
  |||||
  |||||

RESULT 2
ID AAY84264
AC AAY84264;
XX
XX 12-JUL-2000 (first entry)
XX
XX Peptide of alternate reading frame of macrophage colony stimulating gene.
XX
XX Renal cell carcinoma; antigen; cytotoxic T lymphocyte;
XX tumour rejection antigen; macrophage colony stimulating gene;
XX macrophage-colony stimulating factor; antigen presenting cell;
XX human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
XX Homo sapiens.
XX
XX WO200013699-A1.
XX
XX 16-MAR-2000.
XX
XX 03-SEP-1999; 99WO-US020344.
XX
XX 04-SEP-1999; 99WO-US020344.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
XX WPI; 2000-256859/22.
XX
XX Isolated polypeptide used to treat subjects having a disorder
XX characterized by expression of alternative open reading frame macrophage-
XX colony stimulating factor comprises 25 amino acid residue sequence.
XX
XX Example 1; Page 64; 74pp; English.
XX
XX The present sequence represents a tumour rejection antigen precursor, and
XX is encoded by an alternative open reading frame (ORF) of human macrophage
XX colony stimulating gene. Peptides derived from the alternative ORF of
XX macrophage-colony stimulating factor, when presented by an antigen
XX presenting cell having a human leukocyte antigen (HLA) class I molecule,
XX effectively induce the activation and proliferation of CD8+ cytotoxic T
XX lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF
XX of macrophage-colony stimulating factor are useful for enriching
XX selectively a population of T lymphocytes with CD8+ T lymphocytes. They
XX are also useful for diagnosing a disorder characterized by expression of
XX the polypeptide, and for identifying functional variants and mimetics
XX
XX Sequence 25 AA;
XX
XX Query Match      100.0%; Score 106; DB 3; Length 25;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-08;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGLPAVVGSLSPGQEQYHRG 20
Db 1 MAGLPAVVGSLSPGQEQYHRG 20
  |||||
  |||||

RESULT 3
ID AAY84269
AC AAY84269;
XX
XX 12-JUL-2000 (first entry)
XX
XX Peptide derived from macrophage colony stimulating gene alternative ORF.
XX tumour rejection antigen; macrophage colony stimulating gene;
XX macrophage-colony stimulating factor; antigen presenting cell;
XX human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
XX Synthetic.
XX
XX Homo sapiens.
XX
XX WO200013699-A1.
XX
XX 16-MAR-2000.
XX
XX 03-SEP-1999; 99WO-US020344.
XX
XX 04-SEP-1999; 98US-0099077P.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
XX WPI; 2000-256859/22.
XX
XX Isolated polypeptide used to treat subjects having a disorder
XX characterized by expression of alternative open reading frame macrophage-
XX colony stimulating factor comprises 25 amino acid residue sequence.
XX
XX Example 2; Page 40; 74pp; English.
XX
XX The present sequence represents a peptide which is derived from a tumour
XX rejection antigen precursor encoded by an alternative open reading frame
XX (ORF) of human macrophage colony stimulating gene. Peptides derived from
XX the alternative ORF of macrophage-colony stimulating factor, when
XX presented by an antigen presenting cell having a human leukocyte antigen
XX (HLA) class I molecule, effectively induce the activation and
XX proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
XX acids derived from the alternate ORF of macrophage-colony stimulating
XX factor are useful for enriching selectively a population of T lymphocytes
XX with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
XX characterized by expression of the polypeptide, and for identifying
XX functional variants and mimetics
XX
XX Sequence 15 AA;
XX
XX Query Match      70.8%; Score 75; DB 3; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 0.00026;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGLPAVVGSLSPGQEQE 16
Db 1 AGLPAVVGSLSPGQEQE 15
  |||||
  |||||

RESULT 4
ID AAY84266
AC AAY84266;
XX
XX 12-JUL-2000 (first entry)
XX
XX Peptide derived from macrophage colony stimulating gene alternative ORF.
XX tumour rejection antigen; macrophage colony stimulating gene;

```


PS Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour

CC rejection antigen precursor encoded by an alternative open reading frame

CC (ORF) of human macrophage colony stimulating gene. Peptides derived from

CC the alternative ORF of macrophage-colony stimulating factor, when

CC presented by an antigen presenting cell having a human leukocyte antigen

CC (HLA) class I molecule, effectively induce the activation and

CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic

CC acids derived from the alternate ORF of macrophage-colony stimulating

CC factor are useful for enriching selectively a population of T lymphocytes

CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder

CC characterized by expression of the polypeptide, and for identifying

CC functional variants and mimetics

XX

SQ Sequence 13 AA;

Query Match 61.3%; Score 65; DB 3; Length 13;

Best Local Similarity 100.0%; Pred. No. 0.007;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LPAVVGLSPGGEQ 16

DB 1 LPAVVGLSPGGEQ 13

|||||

RESULT 7

ABU49441

ID ABU49441 standard; protein; 277 AA.

XX

AC ABU49441;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #34968.

XX

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

XX Vibrio cholerae.

XX

XX WO200277183-A2.

XX

XX 03-OCT-2002.

XX

XX 21-MAR-2002; 2002WO-US009107.

XX

XX 21-MAR-2001; 2001US-00815242.

XX

XX 06-SEP-2001; 2001US-00948993.

XX

XX 25-OCT-2001; 2001US-0342923P.

XX

XX 08-FEB-2002; 2002US-00072851.

XX

XX 06-MAR-2002; 2002US-0362699P.

XX

XX (ELIT-) ELITRA PHARM INC.

XX

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

XX WPI; 2003-029926/02.

XX

XX N-PSDB; ACA53311.

XX

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX

XX Claim 25; SEQ ID NO 77365; 1765pp; English.

XX

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC on a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 277 AA;

Query Match 47.2%; Score 50; DB 6; Length 277;

Best Local Similarity 52.9%; Pred. No. 33;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 LPAVVGLSPGGEQBYHRG 20

DB 190 LPAFIGAKPYQQAYERG 206

|||||

RESULT 8

AAB69296

ID AAB69296 standard; protein; 191 AA.

XX

AC AAB69296;

XX

DT 12-SEP-2003 (revised)

DT 20-APR-2001 (first entry)

XX

XX HIV-1 non-subtype B clone 94CY032-3 vif protein.

XX

XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpr;

XX vif; vpr; tat; rev; nef; vaccine.

XX

XX Human immunodeficiency virus 1.

XX

XX WO200026416-A1.

XX

XX 11-MAY-2000.

XX

XX 25-OCT-1999; 99WO-US024837.

XX

XX 02-NOV-1998; 98US-00184418.

XX

XX (UABR-) UAB RES FOUND.

XX

XX Hahn BH, Shaw GM, Gao F;

XX

XX WPI; 2000-365651/31.

XX

XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

XX type 1 useful for detecting and treating AIDS comprises a specific

XX nucleotide sequence.

XX

XX Claim 41; Fig 16; 131pp; English.

XX

XX The present in invention provides the protein and coding sequences for a

```
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC isolates. The sequences shown include the near full-length coding
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC rev and nef proteins. These can be used to detect the presence of HIV-1
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC These antibodies can be used in vaccines to prevent and treat HIV
CC infection. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 191 AA;
    Query Match          46.2%; Score 49; DB 3; Length 191;
    Best Local Similarity 66.7%; Pred. No. 31;
    Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
    QY 9 GLSPGEQYHRG 20
    Db 70 GLQPGQDWHLG 81
    RESULT 9
    ID ABU30626 standard; protein; 278 AA.
    XX
    AC ABU30626;
    DT 19-JUN-2003 (first entry)
    DE Protein encoded by Prokaryotic essential gene #16153.
    KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
    OS Haemophilus influenzae.
    XX
    FN WO200277183-A2.
    PD 03-OCT-2002.
    XX
    PF 21-MAR-2002; 2002WO-US009107.
    XX
    PR 21-MAR-2001; 2001US-00815242.
    PR 06-SEP-2001; 2001US-00948993.
    PR 25-OCT-2001; 2001US-0342923P.
    PR 08-FEB-2002; 2002US-00072851.
    PR 06-MAR-2002; 2002US-0362699P.
    XX
    PA (ELIT-) ELITRA PHARM INC.
    PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
    PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
    XX
    DR WPI; 2003-029926/02.
    DR N-PSDB; ACA34496.
    XX
    PT New antisense nucleic acids, useful for identifying proteins or screening
    PT for homologous nucleic acids required for cellular proliferation to
    PT isolate candidate molecules for rational drug discovery programs.
    XX
    PS Claim 25; SEQ ID NO 58550; 1766pp; English.
    XX
    CC The invention relates to an isolated nucleic acid comprising any one of
    CC the 6213 antisense sequences given in the specification where expression
    CC of the nucleic acid inhibits proliferation of a cell. Also included are:
    CC (1) a vector comprising a promoter operably linked to the nucleic acid
    CC encoding a polypeptide whose expression is inhibited by the antisense
    CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
    CC polypeptide or its fragment whose expression is inhibited by the
    CC antisense nucleic acid; (4) an antibody capable of specifically binding
    CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
    CC proliferation or the activity of a gene in an operon required for
    CC proliferation; (7) identifying a compound that influences the activity of
    CC the gene product or that has an activity against a biological pathway
    CC required for proliferation, or that inhibits cellular proliferation; (8)
    CC identifying a gene required for cellular proliferation or the biological
```

CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX SQ Sequence 278 AA;

Query Match 45.3%; Score 48; DB 6; Length 278;
 Best Local Similarity 52.9%; Pred. No. 65;
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 QY 4 LPAAVGLSPGEQYHHRG 20
 ||||| : : : : :
 DB 191 LPAAFGAKPYHQAYRGG 207

RESULT 11
 ABU17217
 ID ABU17217 standard; protein; 287 AA.
 XX
 AC ABU17217;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #2744.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX OS Acinetobacter baumannii.
 XX PN WO200277183-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002WO-US009107.
 XX PR 21-MAR-2001; 2001US-00815242.
 XX PR 06-SEP-2001; 2001US-00948993.
 XX PR 25-OCT-2001; 2001US-0342923P.
 XX PR 08-FEB-2002; 2002US-00072851.
 XX PR 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HJ;
 XX WPI; 2003-029926/02.
 XX DR N-PSDB; ACA21087.
 XX

PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 45141; 1766pp; English.

XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 287 AA;

Query Match 45.3%; Score 48; DB 6; Length 287;
 Best Local Similarity 52.9%; Pred. No. 68;
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 4 LPAAVGLSPGEQYHHRG 20
 ||||| : : : : :
 DB 200 LPAAFGAKPYHQAYRGG 216
 RESULT 12
 ADA36829
 ID ADA36829 standard; protein; 303 AA.
 XX
 AC ADA36829;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Acinetobacter baumannii protein #3990.
 XX
 KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
 XX plant biocontrol agent.
 XX OS Acinetobacter baumannii.
 XX PN US6562958-B1.
 XX PD 13-MAY-2003.
 XX PF 04-JUN-1999; 99US-00328352.
 XX PR 09-JUN-1998; 98US-0088701P.
 XX (GENO-) GENOME THERAPEUTICS CORP.

```

XX Breton G, Bush D;
XX WPI; 2003-576092/54.
XX N-PSDB; ADA32703.
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
XX for diagnosing a bacterial disease, as components of antibacterial
XX vaccines, as targets for antibacterial drugs, or as biocontrol agents for
XX plants.
XX Example; SEQ ID NO 8116; 328pp; English.
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX The A. baumannii nucleic acids and polypeptides are useful as reagents
XX for diagnosing a bacterial disease, as components of antibacterial
XX vaccines, as targets for antibacterial drugs, to detect the presence of
XX A. baumannii and other Acinetobacter species in a sample, in screening
XX compounds for the ability to interfere with the A. baumannii life cycle
XX or to inhibit A. baumannii infection, and as biocontrol agents for
XX plants. The present sequence represents the amino acid sequence of an A.
XX baumannii protein.
XX SQ Sequence 303 AA;

Query Match 45.3%; Score 48; DB 6; Length 303;
Best Local Similarity 52.9%; Pred. No. 72;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 LPAVGLSPGEQBYHRG 20
DB 216 LPAFVGANPYQAYEKG 232

RESULT 13
ABU50765
ID ABU50765 standard; protein; 151 AA.
XX AC ABU50765;
XX DT 07-MAY-2003 (first entry)
XX DE Helicobacter pylori selected interacting domain (SID) protein #108.
XX KW Protein-protein interaction; ulcer; selected interacting domain; SID.
XX OS Helicobacter pylori.
XX PN WO200266501-A2.
XX PD 29-AUG-2002.
XX PF 28-DEC-2001; 2001WO-EP015428.
XX PR 02-JAN-2001; 2001US-0259302P.
XX PA (HYBR-) HYBRIGENICS.
XX PI (INSP ) INST PASTEUR.
XX PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX WPI; 2002-674910/72.
XX N-PSDB; ABX65509.
XX New complexes of protein-protein interactions in Helicobacter pylori,
XX useful for identifying modulating compounds for treating or preventing
XX ulcers in mammals.
XX Claim 6; Page 111; 642pp; English.
XX The invention describes a complex of protein-protein interactions in
XX Helicobacter pylori selected from 421 complexes given in the
XX specification. The complex of protein-protein interactions are useful for
XX screening for agents which modulate the interaction of proteins.
XX Modulating compounds which binds to a targeted bacterial protein may be
XX used for treating or preventing ulcers in a human or animal. This is the
XX amino acid sequence of a selected interacting domain (SID), identified
XX via protein-protein interactions. Note: Where the patent number printed
XX at the top of the pages in the specification has obscured areas of
XX protein sequence, the indexer has replaced the residue with an X to
XX represent an illegible residue
XX SQ Sequence 151 AA;

Query Match 44.3%; Score 47; DB 5; Length 151;
Best Local Similarity 47.1%; Pred. No. 48;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 LPAVGLSPGEQBYHRG 20
DB 79 LPAFVGANPYQAFERG 95

RESULT 14
ABU51750
ID ABU51750 standard; protein; 155 AA.
XX AC ABU51750;
XX DT 07-MAY-2003 (first entry)
XX DE Helicobacter pylori selected interacting domain (SID) protein #1094.
XX KW Protein-protein interaction; ulcer; selected interacting domain; SID.
XX OS Helicobacter pylori.
XX PN WO200266501-A2.
XX PD 29-AUG-2002.
XX PF 28-DEC-2001; 2001WO-EP015428.
XX PR 02-JAN-2001; 2001US-0259302P.
XX PA (HYBR-) HYBRIGENICS.
XX PI (INSP ) INST PASTEUR.
XX PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX WPI; 2002-674910/72.
XX N-PSDB; ABX66495.
XX New complexes of protein-protein interactions in Helicobacter pylori,
XX useful for identifying modulating compounds for treating or preventing
XX ulcers in mammals.
XX Claim 6; Page 354; 642pp; English.
XX The invention describes a complex of protein-protein interactions in
XX Helicobacter pylori selected from 421 complexes given in the
XX specification. The complex of protein-protein interactions are useful for
XX screening for agents which modulate the interaction of proteins.
XX Modulating compounds which binds to a targeted bacterial protein may be
XX used for treating or preventing ulcers in a human or animal. This is the
XX amino acid sequence of a selected interacting domain (SID), identified
XX via protein-protein interactions. Note: Where the patent number printed
XX at the top of the pages in the specification has obscured areas of
XX protein sequence, the indexer has replaced the residue with an X to
XX represent an illegible residue
XX SQ Sequence 155 AA;

Query Match 44.3%; Score 47; DB 5; Length 155;
Best Local Similarity 47.1%; Pred. No. 49;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

CC screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein may be
 CC used for treating or preventing ulcers in a human or animal. This is the
 CC amino acid sequence of a selected interacting domain (SID), identified
 CC via protein-protein interactions. Note: Where the patent number printed
 CC at the top of the pages in the specification has obscured areas of
 CC protein sequence, the indexer has replaced the residue with an X to
 CC represent an illegible residue
 XX SQ Sequence 151 AA;

Query Match 44.3%; Score 47; DB 5; Length 151;
 Best Local Similarity 47.1%; Pred. No. 48;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 4 LPAVGLSPGEQBYHRG 20
 DB 79 LPAFVGANPYQAFERG 95

RESULT 14
 ABU51750
 ID ABU51750 standard; protein; 155 AA.

XX AC ABU51750;
 XX DT 07-MAY-2003 (first entry)
 XX DE Helicobacter pylori selected interacting domain (SID) protein #1094.
 XX KW Protein-protein interaction; ulcer; selected interacting domain; SID.
 XX OS Helicobacter pylori.
 XX PN WO200266501-A2.
 XX PD 29-AUG-2002.
 XX PF 28-DEC-2001; 2001WO-EP015428.
 XX PR 02-JAN-2001; 2001US-0259302P.
 XX PA (HYBR-) HYBRIGENICS.
 XX PI (INSP) INST PASTEUR.
 XX PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
 XX WPI; 2002-674910/72.
 XX N-PSDB; ABX66495.

XX New complexes of protein-protein interactions in Helicobacter pylori,
 XX useful for identifying modulating compounds for treating or preventing
 XX ulcers in mammals.

XX Claim 6; Page 354; 642pp; English.

XX The invention describes a complex of protein-protein interactions in
 XX Helicobacter pylori selected from 421 complexes given in the
 XX specification. The complex of protein-protein interactions are useful for
 XX screening for agents which modulate the interaction of proteins.
 XX Modulating compounds which binds to a targeted bacterial protein may be
 XX used for treating or preventing ulcers in a human or animal. This is the
 XX amino acid sequence of a selected interacting domain (SID), identified
 XX via protein-protein interactions. Note: Where the patent number printed
 XX at the top of the pages in the specification has obscured areas of
 XX protein sequence, the indexer has replaced the residue with an X to
 XX represent an illegible residue

XX SQ Sequence 155 AA;

Query Match 44.3%; Score 47; DB 5; Length 155;
 Best Local Similarity 47.1%; Pred. No. 49;
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

XX The invention describes a complex of protein-protein interactions in
 XX Helicobacter pylori selected from 421 complexes given in the
 XX specification. The complex of protein-protein interactions are useful for

QY 4 LPAVGLSPGEQYHRG 20
Db |||:|:|:|:|:|:|
80 LPAFIGANPYQQAIFRG 96

RESULT 15
AAB69298
ID AAB69298 standard; protein; 192 AA.
XX AC AAB69298;
XX DT 12-SEP-2003 (revised)
DT 20-APR-2001 (first entry)
XX DE HIV-1 non-subtype B clone 962M751-3 vif protein.
XX DE HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpr;
KW vif; vpr; tat; rev; nef; vaccine.
XX OS Human immunodeficiency virus 1.
XX PN WO200026416-A1.
XX PD 11-MAY-2000.
XX PF 25-OCT-1999; 99WO-US024837.
XX PR 02-NOV-1998; 98US-00184418.
XX PA (UABR-) UAB RES FOUND.
XX PI Hahn BH, Shaw GM, Gao F;
XX DR WPI; 2000-365651/31.
XX PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus
PT type 1 useful for detecting and treating AIDS comprises a specific
PT nucleotide sequence.
XX PS Claim 41; Fig 16; 131pp; English.
XX CC The present invention provides the protein and coding sequences for a
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B
CC isolates. The sequences shown include the near full-length coding
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,
CC rev and nef proteins. These can be used to detect the presence of HIV-1
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.
CC These antibodies can be used in vaccines to prevent and treat HIV
CC infection. (Updated on 12-SEP-2003 to standardise OS field)
XX SQ Sequence 192 AA;
Query Match 44.3%; Score 47; DB 3; Length 192;
Best Local Similarity 66.7%; Pred. NO. 62;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 9 GLSPGEQYHRG 20
Db |||:|:|:|:|:|:|
71 GLHFGEREWHUG 82

Search completed: May 7, 2004, 12:33:43
Job time : 56.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:27:20 ; Search time 8.96 Seconds
(without alignments)
150.299 Million cell updates/sec

Title: US-09-786-214A-12
Perfect score: 72
Sequence: 1 LPAVVGLSPGRQEQY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	61.1	381	2 F75270	cytochrome P450 -
2	43	59.7	211	2 A64490	precorrin-6Y methy
3	43	59.7	623	2 T40391	probable lysophosp
4	43	59.7	821	2 C84304	DNA helicase limpo
5	41	56.9	156	2 F87551	conserved hypochet
6	41	56.9	326	2 T45226	probable NS,N10-me
7	41	56.9	508	2 E70764	probable cobi prot
8	40	55.6	227	2 B90400	hypothetical prote
9	40	55.6	243	2 I54459	MEC H-2K1-k - mous
10	40	55.6	428	2 AG1304	uracil permease ho
11	40	55.6	428	2 AG1676	uracil permease ho
12	40	55.6	540	2 A75250	carboxylesterase,
13	40	55.6	661	2 G84511	hypothetical prote
14	40	55.6	673	2 T50281	probable lysophosp
15	39	54.2	43	2 S21065	Ig kappa chain v r
16	39	54.2	96	2 S45441	Ig kappa chain v r
17	39	54.2	103	2 S19975	Ig kappa chain v r
18	39	54.2	106	2 PS0070	Ig kappa chain v r
19	39	54.2	106	2 PC4282	Ig kappa chain (an
20	39	54.2	107	2 S57444	Ig kappa chain v r
21	39	54.2	108	2 C30502	Ig kappa chain v r
22	39	54.2	108	2 S33988	Ig kappa chain v r
23	39	54.2	108	2 G44151	Ig kappa chain v r
24	39	54.2	111	2 S23628	Ig kappa chain v r
25	39	54.2	114	2 S54905	Ig kappa chain v r
26	39	54.2	115	1 K3HUVG	Ig kappa chain pre
27	39	54.2	115	1 KYMSU7	Ig kappa chain pre
28	39	54.2	115	2 S11697	Ig kappa chain pre
29	39	54.2	116	2 B25521	Ig kappa chain pre

30 39 54.2 119 2 S41816
31 39 54.2 125 2 S40344
32 39 54.2 128 2 PN0445
33 39 54.2 128 2 S40379
34 39 54.2 128 2 A56701
35 39 54.2 129 2 S29627
36 39 54.2 129 2 S40363
37 39 54.2 132 2 S05288
38 39 54.2 144 2 P0106
39 39 54.2 144 2 B30502
40 39 54.2 215 2 A23746
41 39 54.2 277 2 P82130
42 39 54.2 278 2 E64131
43 39 54.2 319 2 AD0941
44 39 54.2 563 2 JQ0623
45 39 54.2 1367 2 S74285

ALIGNMENTS

RESULT 1

F75270

cytochrome P450 - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: F75270

R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: F75270

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-381 <WHI>

A:Cross-references: GB:AE002076; GB:AE000513; NID:g6460285; PIDN:AAF12016.1; PID:g64602;

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR2473

A:Map position: 1

Query Match 61.1%; Score 44; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. NO. 12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAVVGLSP 9

Db 52 LPAVVGLSP 60

RESULT 2

A64490

precorrin-6Y methylase homolog - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 18-Aug-2003

C:Accession: A64490

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.

A:Reference number: A64300; MUID:96337999; PMID:8688087

A:Accession: A64490

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-211 <BUL>

A:Cross-references: GB:U67593; GB:L77117; NID:g2826427; PIDN:AAB99541.1; PID:g1592152;

C:Genetics:

A:Map position: FOR1500322-1500957

C:Superfamily: precorrin-6Y methylase CbiE

```

Query Match      59.7%; Score 43; DB 2; Length 211;
Best Local Similarity 54.5%; Pred. No. 9.3;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVGLSPGGEQY 14
   |||: |||: |||
DB 4 IVGIGPGDREY 14

RESULT 3
T40991
probable lysophospholipase precursor - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T40991
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21962
A:Accession: T40991
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-623 <LYN>
A:Cross-references: EMBL:AL049559; PIDN:CA840176.1; GSPDB:GN00068; SPDB:SPCC1450.09c
A:Experimental source: strain 972h-; cosmid ci450
C:Genetics:
A:Gene: SPDB:SPCC1450.09c
A:Map position: 3
C:Superfamily: yeast lysophospholipase

Query Match      59.7%; Score 43; DB 2; Length 623;
Best Local Similarity 69.2%; Pred. No. 29;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PAVVGLSPGGEQY 14
   || ||| ||| |||
DB 76 PASDGLSTGEQRF 88

RESULT 4
C84304
DNA helicase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84304
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: C84304
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-821 <STO>
A:Cross-references: GB:AE004437; NID:gl0580995; PIDN:AAG19799.1; GSPDB:GN00138
C:Genetics:
A:Gene: hel

Query Match      59.7%; Score 43; DB 2; Length 821;
Best Local Similarity 90.0%; Pred. No. 38;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AVVGLSPGGEQ 12
   ||| ||| ||| |||
DB 326 AVVGLSPAEG 335

RESULT 5
F87551
conserved hypothetical protein CC2439 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus

```

```

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: F87551
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87551
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <STO>
A:Cross-references: GB:AE005673; NID:gl3423984; PIDN:AAK24410.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2439
C:Superfamily: Haemophilus influenzae conserved hypothetical protein HI0305

Query Match      56.9%; Score 41; DB 2; Length 156;
Best Local Similarity 80.0%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 AVVGLSPGGEQ 12
   ||| ||| |||
DB 18 AVVGLDPGEK 27

RESULT 6
T45226
probable NS,N10-methylene-tetrahydromethanopterin reductase (F420-dependent) [imported]
C:Species: Methanobolus tindarius
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45226
R:Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G.;
submitted to the EMBL Data Library, September 1998
A:Description: The F420H2-dehydrogenase from Methanobolus tindarius: Cloning of the ffa
A:Reference number: Z22947
A:Accession: T45226
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-326 <WES>
A:Cross-references: EMBL:AJ011519; PIDN:CAB56639.1
A:Experimental source: DSM 2278
C:Genetics:
A:Gene: ffa

Query Match      56.9%; Score 41; DB 2; Length 326;
Best Local Similarity 70.0%; Pred. No. 32;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 AVVGLSPGGEQ 12
   ||| ||| ||| |||
DB 88 AILGLPGGEQ 97

RESULT 7
E70764
probable cbl protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70764
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70764
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-508 <COL>
A:Cross-references: GB:Z73966; GB:AL123456; NID:g3261577; PIDN:CAA98214.1; PID:e246996;
A:Experimental source: strain H37RV

```

C;Genetics:
A;Gene: cobi

Query Match 56.9%; Score 41; DB 2; Length 508;
Best Local Similarity 58.3%; Pred. No. 51;
Matches 7; Conservative 3; Mismatches 0; Gaps 0;

Qy 3 AVVGLSPGEOEY 14
Db 250 AVVGLSPGDSNW 261

RESULT 8

B90400
hypothetical protein cbiE [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 18-Aug-2003
C;Accession: B90400
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrest, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: B90400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-227 <KUR>
A;Cross-references: GB:AE006641; NID:gl3815599; PIDN:AAK42457.1; GSPDB:GN00155
C;Genetics:
A;Gene: cbiE
C;Superfamily: precorrin-6Y methylase CbiE

Query Match 55.6%; Score 40; DB 2; Length 227;
Best Local Similarity 54.5%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 0; Gaps 0;

Qy 4 VVGLSPGEOEY 14
Db 10 IVGVGPGDEY 20

RESULT 9

I54459
MHC H-2K1-k - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I54459
R;Watts, S.; Davis, A.C.; Goodenow, R.S.
Immunogenetics 29, 355-357, 1989
A;Title: Sequence analysis of the C3H H-2K1-k gene: Relationship to the H-2 loci.
A;Reference number: I54459; MUID:89233303; PMID:2714856
A;Accession: I54459
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-243 <RES>
A;Cross-references: GB:M27134; NID:gl99435; PIDN:AAA39610.1; PID:g387456
C;Genetics:
A;Introns: 22/1; 112/1

Query Match 55.6%; Score 40; DB 2; Length 243;
Best Local Similarity 77.8%; Pred. No. 35;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VGLSPGEOE 13
Db 216 LGLSPGEE 224

RESULT 10

AG1304
uracil permease homolog pyrP [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes

C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AG1304

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke-
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mi-
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AG1304

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-428 <GLA>

A;Cross-references: GB:NC_003210; PIDN:CAC99917.1; PID:gl6411293; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: pyrP

C;Superfamily: uracil transport protein uraA

Query Match 55.6%; Score 40; DB 2; Length 428;
Best Local Similarity 70.0%; Pred. No. 63;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPG 10
Db 42 VPSVTGLSPG 51

RESULT 11

AG1676

uracil permease homolog pyrP [imported] - Listeria innocua (strain Clip11262)

C;Species: Listeria innocua

C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C;Accession: AG1676

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke-
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M-
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A;Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AG1676

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-428 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC97183.1; PID:gl6414454; GSPDB:GN00178

A;Experimental source: strain Clip11262

C;Genetics:

A;Gene: pyrP

C;Superfamily: uracil transport protein uraA

Query Match 55.6%; Score 40; DB 2; Length 428;
Best Local Similarity 70.0%; Pred. No. 63;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPG 10
Db 42 VPSVTGLSPG 51

RESULT 12

A75250

carboxylesterase, type B - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C;Accession: A75250

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma-
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: A75250
 A;Molecule type: DNA
 A;Residues: 1-540 <WHI>
 A;Cross-references: GB:AE002092; GB:AE000513; NID:g6460455; PIDN:AAF12163.1; PID:g646045
 A;Experimental source: strain R1
 C;Genetics:

A;Map position: 1
 A;Gene: DR2626
 C;Superfamily: cholinesterase; cholinesterase homology

Query Match 55.6%; Score 40; DB 2; Length 540;
 Best Local Similarity 70.0%; Pred. No. 80;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PAVVGLSPGE 11
 ||:|||||
 Db 512 PQVGLAPGE 521

RESULT 13

GB4511

hypothetical protein At2g13900 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C;Accession: G84511

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Xoo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
 eus, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: G84511

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-661 <STO>

A;Cross-references: GB:AE002093; NID:g6598598; PIDN:AAF18650.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g13900

A;Map position: 2

Query Match 55.6%; Score 40; DB 2; Length 661;
 Best Local Similarity 53.8%; Pred. No. 99;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQY 14
 ||:|||||
 Db 437 PTLKIVPGEQY 449

RESULT 14

T50281

probable lysophospholipase (EC 3.1.1.5) precursor SPAC977.09c [similarity] - fission yea

C;Species: Schizosaccharomyces pombe

C;Date: 03-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 19-Jan-2001

C;Accession: T50281; T42738

R;Zimmermann, W.; Wambutt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, January 2000

A;Reference number: Z25053

A;Accession: T50281

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-673 <ZIM>

A;Cross-references: EMBL:AL1317130; NID:g6742151; PIDN:CAB69631.1; PID:g6742159; GSPDB:GN

A;Experimental source: strain 972h(-); Cosmid c977

R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.

DNA Res. 4, 363-369, 1997

A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.

A;Reference number: Z17323; MUID:98162722; PMID:9501991

A;Accession: T42738

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 427-673 <YOS>

A;Cross-references: EMBL:D89183; NID:gl749573; PIDN:BAA13845.1; PID:gl749574
 A;Experimental source: strain PR745

C;Genetics:

A;Gene: SPDB:SPAC977.09c

A;Map position: 1

A;Introns: 651/3

C;Function:

A;Description: catalyzes the hydrolysis of 2-lysophosphatidylcholine to glycerophosphoch
 C;Superfamily: yeast lysophospholipase
 C;Keywords: carboxylic ester hydrolase

Query Match 55.6%; Score 40; DB 2; Length 673;
 Best Local Similarity 61.5%; Pred. No. 1e+02;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQY 14
 ||:|||||
 Db 83 PASEGLNEGEQY 95

RESULT 15

S21065

Ig kappa chain V region (anti-RH(D)) - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 22-Nov-1993 #sequence_revision 24-May-1996 #text_change 09-May-1997

C;Accession: S21065

R;Dlouha, A.; Lecroisey, A.; Henschen, A.; Rouger, P.; Keil, B.

Protein Seq. Data Anal. 4, 317-318, 1991

A;Title: Subgroup assignment of a human monoclonal anti-Rh(D) antibody.

A;Reference number: S21065; MUID:92253544; PMID:1812483

A;Accession: S21065

A;Molecule type: protein

A;Residues: 1-43 <DLO>

C;Superfamily: immunoglobulin V region; immunoglobulin homology

C;Keywords: heterotetramer; immunoglobulin

Query Match 54.2%; Score 39; DB 2; Length 43;
 Best Local Similarity 63.6%; Pred. No. 8.4;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
 ||:|||||
 Db 8 PATLSLSPGE 18

Search completed: May 7, 2004, 12:39:06
 Job time : 9.12667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 5.04 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-12

Perfect score: 72

Sequence: 1 LPAVGLSPGEQY 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	59.7	211	1	COBL_METJA
2	43	59.7	633	1	PLB5 SCHPO
3	41	56.9	156	1	RUVX CAUCR
4	41	56.9	326	1	VER METRI
5	41	56.9	508	1	COBI MYCTU
6	40	55.6	673	1	PLB4 SCHPO
7	39	54.2	115	1	KV31 HUMAN
8	39	54.2	115	1	KV51 MOUSE
9	39	54.2	278	1	PURU HAEIN
10	39	54.2	516	1	C4AD DROME
11	39	54.2	597	1	NR41 RAT
12	39	54.2	1402	1	N160 MOUSE
13	39	54.2	1636	1	BUD3 YEAST
14	38	52.8	390	1	COBL MYCTU
15	38	52.8	429	1	RS1 LEULA
16	38	52.8	507	1	CAT4 PICAN
17	38	52.8	699	1	EFG HAEIN
18	38	52.8	700	1	EFG PASMO
19	37.5	52.1	827	1	MAK1 MOUSE
20	37	51.4	280	1	PMXA MOUSE
21	37	51.4	281	1	PMXA RAT
22	37	51.4	446	1	COBJ ARCFU
23	37	51.4	446	1	ENO1 MAIZE
24	37	51.4	557	1	PUR6 VIGAC
25	37	51.4	607	1	GLMS CLOTE
26	37	51.4	637	1	MUTL CAUCR
27	37	51.4	658	1	VG18 BPT4
28	37	51.4	753	1	CKXA BACUF
29	37	51.4	813	1	CADM_MOUSE
30	37	51.4	985	1	4ET HUMAN
31	37	51.4	992	1	EVC HUMAN
32	37	51.4	2269	1	RRB1 SV41
33	36.5	50.7	374	1	RGSK_BOVIN

34	36.5	50.7	1121	1	BMS1 SCHPO
35	36	50.0	129	1	KV3H HUMAN
36	36	50.0	158	1	RSD_ECOLI
37	36	50.0	162	1	RSD_SALTY
38	36	50.0	243	1	SUMT SYN7
39	36	50.0	279	1	TYSY CAUCR
40	36	50.0	280	1	PURU ECOL6
41	36	50.0	280	1	PURU ECOLI
42	36	50.0	317	1	OADI_HUMAN
43	36	50.0	342	1	HUPK_AZOVI
44	36	50.0	348	1	HOXV_AZOVI
45	36	50.0	359	1	ALF2_PEA

ALIGNMENTS

RESULT 1

COBL_METJA	STANDARD;	PRT;	211 AA.
AC	Q58917;	1998 (Rel. 37, Created)	
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Probable precorrin-6Y C5,15-methyltransferase [decarboxylating]		
DE	(EC 2.1.1.132) (Precorrin-6 methyltransferase) (Precorrin-6Y		
DE	methylase)		
GN	COBL OR MJ1522.		
OS	Methanococcus jannaschii.		
OC	Archaea; Euryarchaeota; Methanococci; Methanococcales;		
OC	Methanocaldococcaceae; Methanocaldococcus.		
OX	NCBI_TaxID=2190;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;		
RX	MEDLINE=96337999; PubMed=8688087;		
RA	Bult G.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,		
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gayne J.D.,		
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,		
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,		
RA	Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,		
RA	Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,		
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,		
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;		
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus		
RL	jannaschii".		
RL	Science 273:1058-1073 (1996).		
CC	-1- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN		
CC	PRECORRIN-6Y TO FORM PRECORRIN-8X (BY SIMILARITY).		
CC	-1- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2		
CC	S-adenosyl-L-homocysteine + precorrin-8X + CO(2).		
CC	-1- PATHWAY: Cobalamin biosynthesis.		
CC	-1- SIMILARITY: TO S-TYRPHIMURIUM CBIE; ALSO, LOW, TO OTHER		
CC	METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.		

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U67593; AAB93541.1; --
PIR; A64490; A64490.
TIGR; MJ1522; --
InterPro; IPR000878; Cor/por_Mettransf.
Pfam; PF00590; TP_methylase; 1.
Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
Methyltransferase; Complete proteome.
KW
SEQUENCE 211 AA; 23805 MW; 279A1A2B14369510 CRC64;

Query Match 59.7%; Score 43; DB 1; Length 211;

DR EMBL; AE005913; AAK24410.1; --
 DR PIR; F87551; F87551.
 DR TIGR; CC2439; --
 DR HAMAP; MF 00651; -- 1.
 DR InterPro; IPR005227; Cons_hypoth250.
 DR InterPro; IPR006641; YqgFC.
 DR Pfam; PF03652; UFP0081; 1.
 DR SMART; SM00732; YqgFc; 1.
 DR TIGRFAMs; TIGR00250; TIGR00250; 1.
 KW Hydrolyase; Nuclease; DNA repair; DNA recombination; Complete proteome.
 SQ SEQUENCE 156 AA; 17142 MW; 21F54D8648396141 CRC64;

Query Match 56.9%; Score 41; DB 1; Length 156;
 Best Local Similarity 80.0%; Pred. No. 9.4;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 AVVGLSPGEQ 12
 |||||
 Db 18 AVVGLDPGEK 27
 |||||

RESULT 4
 MER_METTI
 ID_MER_METTI STANDARD; PRT; 326 AA.
 AC Q9UXP0;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Coenzyme F420-dependent N(5),N(10)-methylentetrahydromethanopterin
 DE reductase (EC 1.5.99.11) (Methylene-H(4)MPT reductase).
 GN MER OR FFD4.
 OS Methanobolus tindarius.
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
 OC Methanosarcinaceae; Methanobolus.
 OX NCBI_TaxID=2221;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 2278;
 RX MEDLINE=99132696; PubMed=9933933;
 RA Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
 RA Gottschalk G., Blaut M.;
 RT "The F420H2-dehydrogenase from Methanobolus tindarius: cloning of the
 RT ffd operon and expression of the genes in Escherichia coli";
 RL FEMS Microbiol. Lett. 170:389-398(1999).
 CC -!- FUNCTION: Catalyzes the reversible reduction of methylene-H(4)MPT
 CC to methyl-H(4)MPT (By similarity).
 CC -!- CATALYTIC ACTIVITY: N(5),N(10)-methylentetrahydromethanopterin +
 CC reduced coenzyme F420 = 5-methyl-5,6,7,8-tetrahydromethanopterin +
 CC coenzyme F420.
 CC -!- PATHWAY: Methanogenesis from carbon dioxide; fifth step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: Belongs to the mer family.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; AJ011519; CAB56639.1; --
 DR PIR; T45226; T45226.
 DR HAMAP; MF 01091; -- 1.
 DR InterPro; IPR002103; Bac_luciferase.
 DR Pfam; PF00296; bac_luciferase; 1.
 KW Mechanogenesis; One-carbon metabolism; Oxidoreductase.
 SQ SEQUENCE 326 AA; 34043 MW; 16F3AB9733A45DB2 CRC64;

Query Match 56.9%; Score 41; DB 1; Length 326;
 Best Local Similarity 70.0%; Pred. No. 19;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 AVVGLSPGEQ 12
 |||||
 Db 88 AILGLGPGEQ 97
 |||||

RESULT 5
 COBI_MYCTU
 ID_COBI_MYCTU STANDARD; PRT; 508 AA.
 AC Q16677;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cobalamin biosynthesis protein COBIJ [includes: Precorrin-2 C20-
 DE methyltransferase (PC 2.1.1.130) (S-adenosyl-L-methionine-precorrin-2
 DE methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
 DE COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon K.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Oxborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulten J.B., Taylor K., Whitehead S., Barrall B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Mounsepe C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RA "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
 CC PRECORRIN-3A (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
 CC adenosyl-L-homocysteine + precorrin-3A.
 CC -!- PATHWAY: Cobalamin biosynthesis.
 CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
 CC AND CBIL/COBI.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).


```

CC EMBL; Z73966; CAA98214.1; -
DR EMBL; AE007063; KAK46406.1; -
DR EMBL; BX248341; CAD96945.1; -
DR PIR; E70764; E70764.
DR TIGR; WP2126; -
DR Tuberculist; RV2066; -
DR InterPro; IPR006364; Cobi_CbiL.
DR InterPro; IPR006363; Cobi.
DR InterPro; IPR000878; Cor/por Metransf.
DR InterPro; IPR003043; Uropor Metransf.
DR Pfam; PR00590; TP methylase.2.
DR TIGRPFAM; TIGR01457; cobi_cbiL; 1.
DR TIGRPFAM; TIGR01466; cobi_cbiH; 1.
DR PROSITE; PS00839; SUMT_1; 1.
DR PROSITE; PS00840; SUMT_2; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Multifunctional enzyme; Complete proteome.
FT DOMAIN 244 508 PRECORRIN-2 C20-METHYLTRANSFERASE.
FT DOMAIN 244 508 PRECORRIN-3 METHYLASE.
SQ SEQUENCE 508 AA; 53910 MW; 95AC066F02C24DC1 CRC64;

Query Match 56.9%; Score 41; DB 1; Length 508;
Best Local Similarity 58.3%; Pred. No. 30;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 AVVGLSPGQEY 14
||||| ||: ||
Db 250 AVVGLGPGSDW 261

RESULT 6
PLB4_SCHPO STANDARD; PRT; 673 AA.
AC Q9P327;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative lysophospholipase C977.09c precursor (EC 3.1.1.5)
DE (Phospholipase B).
GN SPAC977.09C OR SPAC1348.10C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

```

```

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: Catalyzes the release of fatty acids from
CC lysophospholipids (by similarity).
CC -!- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC glycerophosphocholine + a fatty acid anion.
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- SIMILARITY: Belongs to the lysophospholipase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL358912; CAB94277.1; -
CC EMBL; AL137130; CAB69631.1; -
CC PIR; T50281; T50281.
CC GenDB_SPombe; SPAC977.09c; -.
CC InterPro; IPR001179; FKBP_PPase.
CC InterPro; IPR002642; PLAC.
CC Pfam; PF01735; PLA2_B; 1.
CC SMART; SM0022; PLAC; 1.
KW Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
KW Signal.
FT SIGNAL 1 19
FT CHAIN 20 673 POTENTIAL. LYSOPHOSPHOLIPASE C977.09C.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 516 516 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 574 574 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 673 AA; 74595 MW; B39A773E76CD694B CRC64;

Query Match 55.6%; Score 40; DB 1; Length 673;
Best Local Similarity 61.5%; Pred. No. 58;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAVVGLSPGQEY 14
||| ||: |||
Db 83 PASEGLNEGEQSY 95

RESULT 7
KV31_HUMAN STANDARD; PRT; 115 AA.
ID KV31_HUMAN
AC P04433;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG kappa chain V-JII region VG precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85087932; PubMed=6440122;
RA Pech M., Zachau H.G.;
RA "Immunoglobulin genes of different subgroups are interdigitated
RT within the VK locus.";
RL Nucleic Acids Res. 12:9229-9236(1984).

```

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X01668; -; NOT_ANNOTATED_CDS.
 CC PIR; A01900; K3HUVG.
 CC HSSP; P80362; 1WTL.
 CC GO; GO:0005576; C:extracellular; NAS.
 CC GO; GO:0003823; F:antigen binding; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003596; Ig_v.
 CC Pfam; PF00047; Ig; 1.
 CC SMART; SM00406; IGV; 1.
 CC PROSITE; PS50835; IG_LIKE; 1.
 CC Immunoglobulin V region; Signal.
 KW Immunoglobulin V region; Signal; 3D-structure.
 FT SIGNAL 1 20
 FT CHAIN 21 >115 IG KAPPA CHAIN V-III REGION VG.
 FT DOMAIN 21 43 FRAMEWORK-1.
 FT DOMAIN 44 54 COMPLEMENTARITY-DETERMINING-1.
 FT DOMAIN 55 69 FRAMEWORK-2.
 FT DOMAIN 70 76 COMPLEMENTARITY-DETERMINING-2.
 FT DOMAIN 77 108 FRAMEWORK-3.
 FT DOMAIN 109 115 COMPLEMENTARITY-DETERMINING-3.
 FT DISULFID 43 108 BY SIMILARITY.
 FT NON_TER 115 115
 SQ SEQUENCE 115 AA; 12575 MW; 2DR47CDA3A17D555 CRC64;
 Query Match 54.2%; Score 39; DB 1; Length 115;
 Best Local Similarity 63.6%; Pred. No. 15;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 PAVVGLSPGEQ 12
 DB 28 PAVLSVSPGER 38
 RESULT A
 KV51_MOUSE
 ID KV51_MOUSE STANDARD; PRT; 115 AA.
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ig kappa chain v-v region L7 precursor (Fragment).
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81220975; PubMed=6264318;
 RA Pech M., Hochti J., Schneil H., Zachau H.G.;
 RT "Differences between germ-line and rearranged immunoglobulin v kappa
 RT coding sequences suggest a localized mutation mechanism.";
 RL Nature 291:668-670(1981).
 CC -1- MISCELLANEOUS: THERE APPEAR TO BE TWO POSSIBLE SPICE JUNCTIONS AT
 CC THE 3' END OF THE INTRON. THE ALTERNATE WOULD CODE FOR A PROTEIN
 CC LACKING RESIDUES 17-19.
 CC PIR; A01925; KVM5L7.
 CC PDB; 1J10; 18-FEB-03.
 CC PDB; 1J1P; 18-FEB-03.
 CC PDB; 1J1X; 18-FEB-03.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003596; Ig_v.
 CC Pfam; PF00047; Ig; 1.
 CC SMART; SM00406; IGV; 1.
 CC PROSITE; PS50835; IG_LIKE; 1.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X01668; -; NOT_ANNOTATED_CDS.
 CC PIR; A01900; K3HUVG.
 CC HSSP; P80362; 1WTL.
 CC GO; GO:0005576; C:extracellular; NAS.
 CC GO; GO:0003823; F:antigen binding; NAS.
 CC GO; GO:0006955; P:immune response; NAS.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003596; Ig_v.
 CC Pfam; PF00047; Ig; 1.
 CC SMART; SM00406; IGV; 1.
 CC PROSITE; PS50835; IG_LIKE; 1.
 KW Immunoglobulin V region; Signal; 3D-structure.
 FT SIGNAL 1 20
 FT CHAIN 21 >115 IG KAPPA CHAIN V-V REGION L7.
 FT DOMAIN 21 43 FRAMEWORK-1.
 FT DOMAIN 44 54 COMPLEMENTARITY-DETERMINING-1.
 FT DOMAIN 55 69 FRAMEWORK-2.
 FT DOMAIN 70 76 COMPLEMENTARITY-DETERMINING-2.
 FT DOMAIN 77 108 FRAMEWORK-3.
 FT DOMAIN 109 115 COMPLEMENTARITY-DETERMINING-3.
 FT DISULFID 43 108 BY SIMILARITY.
 FT NON_TER 115 115
 SQ SEQUENCE 115 AA; 12615 MW; C17BEC758C577E00 CRC64;
 Query Match 54.2%; Score 39; DB 1; Length 115;
 Best Local Similarity 54.5%; Pred. No. 15;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 2 PAVVGLSPGEQ 12
 DB 28 PAVLSVSPGER 38
 RESULT 9
 PURU_HAEIN
 ID PURU_HAEIN STANDARD; PRT; 278 AA.
 AC Q03432;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)
 DE hydrolase).
 GN PURU OR H11588.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 CC Pasteurellaceae; Haemophilus.
 CC NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerkelavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 RN [2]
 RP SEQUENCE OF 64-278 FROM N.A.
 RC STRAIN=RM 7004 / Serotype B;
 RX MEDLINE=93328119; PubMed=8335255;
 RA Maskell D.J.;
 RT "Cloning and sequencing of the Haemophilus influenzae ara gene.";
 RL Gene 129:155-156(1993).
 CC -1- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES
 CC THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF
 CC 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC
 CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL
 CC (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +
 CC tetrahydrofolate.
 CC -1- ENZYME REGULATION: Activated by methionine, inhibited by glycine
 CC (By similarity).
 CC -1- PATHWAY: De novo purine biosynthesis.
 CC -1- SUBUNIT: Homohexamer (By similarity).
 CC -1- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U02833; AAC23236.1; -;
DR EMBL: L04686; AAA44942.1; -;
DR PIR: E64131; E64131.
DR HSSP: P08179; 1GRC.
DR TIGR: H11588; -;
DR InterPro: IPR002912; ACT.
DR InterPro: IPR002376; formyl_transf.
DR InterPro: IPR004810; PurU.
DR Pfam: PF01842; ACT; 1.
DR Pfam: PF00551; formyl_transf; 1.
DR PRINTS: PR01575; FFH4HYDLASE.
DR TIGRFAMs: TIGR00655; PurU; 1.
KW Purine biosynthesis; Hydrolase; One-carbon metabolism;
KW Complete proteome.
FT ACT_SITE 223 BY SIMILARITY.
FT CONFLICT 115 117 VIG -> RNR (IN REF. 2).
FT CONFLICT 138 140 HEN -> PK (IN REF. 2).
FT CONFLICT 205 205 K -> E (IN REF. 2).
SQ SEQUENCE 278 AA; 32173 MW; 7F375AB3C422EC4B CRC64;
Query Match 54.2%; Score 39; DB 1; Length 278;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 LPVAVGLSPGEQRY 14
Db 191 LPAFIGKPYQAY 204

RESULT 10
C4AD_DROME STANDARD; PRT; 516 AA.
AC Q9V4T3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable cytochrome P450 4ad1 (EC 1.14.-.-) (CYPTVAD1).
GN CYP4AD1 OR CG2110.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laspo P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
CC -!- FUNCTION: May be involved in the metabolism of insect hormones and
CC in the breakdown of synthetic insecticides (By similarity).
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
CC (Potential).
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AB003837; AAF59092.1; -;
DR EMBL: AY061058; AAL28606.1; -;
DR HSSP: P14779; 1JPZ.
DR FlyBase: FBgn0033292; Cyp4ad1.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum; Hypothetical protein.
FT METAL 445 445 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 516 AA; 58870 MW; 648EA22492AF59C7 CRC64;
Query Match 54.2%; Score 39; DB 1; Length 516;
Best Local Similarity 57.1%; Pred. No. 65;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 LPVAVGLSPGEQRY 14
Db 469 LPVAVGLPPGINDH 482

RESULT 11
NR41 RAT STANDARD; PRT; 597 AA.
AC P22829;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

Orphan nuclear receptor HMR (Nerve growth factor induced protein I-B) (NGFI-B) (NUR77).

NR4A1 OR HMR OR NGFI-B.

Rattus norvegicus (Rat).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. NCBI_TaxID=10116;

SEQUENCE FROM N.A. MEDLINE=90166506; PubMed=3272167;

Milbrandt J.; "Nerve growth factor induces a gene homologous to the glucocorticoid receptor gene.";

Neuron 1:183-188(1986).

[2]

CHARACTERIZATION. MEDLINE=93361012; PubMed=895013;

Wilson T.E., Fahrner T.J., Milbrandt J.;

"The orphan receptors NGFI-B and steroidogenic factor 1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction.";

Mol. Cell. Biol. 13:5794-5804(1993).

[3]

DNA BINDING MOTIFS. MEDLINE=92229411; PubMed=1314418;

Wilson T.E., Paulsen R.E., Padgett K.A., Milbrandt J.;

"Participation of non-zinc finger residues in DNA binding by two nuclear orphan receptors.";

Science 256:107-110(1992).

[4]

PHOSPHORYLATION. MEDLINE=94043340; PubMed=8227042;

Hirata Y., Kiuchi K., Chen H.-C., Milbrandt J., Guroff G.;

"The phosphorylation and DNA binding of the DNA-binding domain of the orphan nuclear receptor NGFI-B.";

J. Biol. Chem. 268:24808-24812(1993).

[5]

X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 264-351 IN COMPLEX WITH NBRE, AND METAL-BINDING. MEDLINE=99260743; PubMed=10331876;

Meinke G., Sigler P.B.;

"DNA-binding mechanism of the monomeric orphan nuclear receptor NGFI-B.";

Nat. Struct. Biol. 6:471-477(1999).

FUNCTION: Probable nuclear receptor. May act concomitantly with NUR1 in regulating the expression of delayed-early genes during liver regeneration. Binds the NGFI-B response element (NBRE) 5'-AAAAGGTCA-3'.

SUBUNIT: Binds DNA as a monomer.

SUBCELLULAR LOCATION: Nuclear.

TISSUE SPECIFICITY: Expressed in lung, brain and superior cervical ganglia. High levels are seen in the adrenal tissue.

INDUCTION: By nerve growth factor and during liver regeneration.

PTM: Phosphorylation of Ser-350 results in decrease in NBRE binding while phosphorylation of Ser-340 has little effect on it.

SIMILARITY: Belongs to the nuclear hormone receptor family. NR4 subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U17254; AAA56770.1; ALT_INIT.

DB; 1C1T; 26-JUN-00.

TRANSFAC; T00619;

InterPro; IPR000536; Hormone_rec_lig.

InterPro; IPR001723; Strhmn_receptor.

InterPro; IPR008946; Str_ncl_receptor.

InterPro; IPR001628; Znf_C4steroid.

PFam; PF00104; hormone_rec; 1.

PFam; PF00105; zf-C4; 1.

PRINTS; PRO0398; STRDHORMONER.

PRINTS; PRO0047; STROIDFINGER.

ProDom; PRO00035; Znf_C4steroid; 1.

SMART; SM00430; HOLI; 1.

SMART; SM00399; Znf_C4; 1.

PROSITE; PS00031; NUCLEAR RECEPTOR; 1.

Receptor; Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger; Phosphorylation; 3D-structure.

FT DNAS_BIND 266 331 NUCLEAR RECEPTOR-TYPE.

FT ZN_FING 266 286 C4-TYPE.

FT ZN_FING 302 326 C4-TYPE.

FT DOMAIN 408 458 LIGAND-BINDING (POTENTIAL).

FT DOMAIN 80 91 POLY-SER.

FT DOMAIN 182 186 POLY-PRO.

FT DOMAIN 582 585 POLY-PRO.

FT MOD_RES 340 340 PHOSPHORYLATION (BY PKA).

FT MOD_RES 350 350 PHOSPHORYLATION (BY PKA).

FT MUTAGEN 340 340 S->A: LOSS OF PHOSPHORYLATION.

FT MUTAGEN 350 350 S->A: LOSS OF PHOSPHORYLATION.

FT MUTAGEN 345 345 R->K: DECREASED NERE BINDING.

FT MUTAGEN 348 348 L->V: ALMOST COMPLETE LOSS OF NERE BINDING.

FT SEQUENCE 597 AA; 64281 MW; 9CFA987112337E53 CRC64;

SQ

Query Match 54.2%; Score 39; DB 1; Length 597;

Best Local Similarity 46.2%; Pred. No. 75;

Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LPVVVGLSPGQEQ 13

Db 423 IPGFIELSPGDQD 435

RESULT 12

NI60 MOUSE

ID NI60 MOUSE STANDARD; PRT; 1402 AA.

AC Q9Z0W3; Q9CZD3;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Nuclear pore complex protein Nup160 (Nucleoporin Nup160) (160 kDa DE nucleoporin) (Gene trap locus 1-13) (GTL-13).

OS NUP160 OR GTL1-13 OR KIAA0197.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

PN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=129/SVJ;

RA Van de Putte T., Cozijneen M., Dewulf N., Tylzanowski P., Lonnoy O., RA Huybreck D.;

RT KIAA0197 gene (DB3781), complete cds.";

RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RL [2]

RN SEQUENCE OF 1151-1402 FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Embryo;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., RA Kadoya K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T., RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H., RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P., RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
 RA Hayaishizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP IDENTIFICATION, SUBUNIT, AND SUBCELLULAR LOCATION.
 RX MEDLINE=21448620; PubMed=11564755;
 RA Belgareh N., Rabut G., Bai S.W., van Overbeek M., Beaudouin J.,
 RA Daigle N., Zatepina O.V., Pasteau F., Labas V., Fromont-Racine M.,
 RA Ellenberg J., Doye V.,
 RT "An evolutionarily conserved NPC subcomplex, which redistributes in
 part to kinetochores in mammalian cells.";
 RL J. Cell Biol. 154:1147-1160(2001).
 RN [4]
 RP IDENTIFICATION, FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
 RX MEDLINE=21541555; PubMed=11684705;
 RA Vasu S., Shah S., Orjalo A., Park M., Fischer W.H., Forbes D.J.,
 RT "Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA
 export.";
 RL J. Cell Biol. 155:339-354(2001).
 CC -!- SUBUNIT: Involved in poly(A) + RNA transport.
 CC -!- SUBUNIT: Forms part of the Nup160 subcomplex in the nuclear pore
 CC which is composed of Nup160, Nup133, Nup107 and Nup96. This
 CC complex plays a role in RNA export and in tethering Nup98 and
 CC Nup153 to the nucleus.
 CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
 CC frameshift in position 1157 and a stop codon in position 1396.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF104415; AAD17922.2;
 CC EMBL; AK012715; BAB28429.1; ALT_FRAME.
 CC GCD; MGI:1926227; Nup160.
 CC GO; GO:0005643; C:nuclear pore; IDA.
 CC GO; GO:0005487; F:nucleocytoplasmic transporter activity; IDA.
 CC GO; GO:0006406; P:mRNA-nucleus export; IDA.
 CC Nuclear protein; Transport.
 CC CONFLICT 1156 1156 A -> T (IN REF. 2).
 CC CONFLICT 1314 1314 E -> G (IN REF. 2).
 CC CONFLICT 1368 1368 N -> D (IN REF. 2).
 CC SEQUENCE 1402 AA; 158230 MW; 3BF5DF057D28772 CRC64;
 Query Match 54.2%; Score 39; DB 1; Length 1402;
 Best local Similarity 70.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LPAVVGSLSPG 10
 Dd 170 IPSVFGSLSPG 179
 RESULT 13
 BUD3 YEAST
 ID_BUD3 YEAST STANDARD; PRT; 1636 AA.
 AC P25556; P25557; P87007;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bud site selection protein BUD3.
 GN BUD3 OR YCL014W/YCL013W/YCL012W OR YCL14W/YCL13W/YCL12W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 [1] SEQUENCE FROM N.A.
 RX MEDLINE=95247824; PubMed=7730410;
 RA Chant J., Mischke M., Mitchell E., Herskowitz I., Pringle J.R.;
 RT "Role of Bud3p in producing the axial budding pattern of yeast.";
 RL J. Cell Biol. 129:767-778(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=92244356; PubMed=1574125;
 RA Oliver S.G.; van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
 RA Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
 RA Bent P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
 RA Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
 RA Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
 RA De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
 RA Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Fabre F.,
 RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
 RA Francinques-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
 RA Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Glansdorff N.,
 RA Goffeau A., Grenson M., Grisanti P., Griwell L.A., Haesebaun M.,
 RA Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
 RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
 RA Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
 RA Kleinhaus U., Kreisel P., Lafranchi G., Lewis C., van der Linden C.G.,
 RA Lucchini G., Kutzschenken K., Maat C., Manhaupt M., Manzano M.E.,
 RA Lucchini G., Kutzschenken K., Maat C., Manhaupt M., Manzano M.E.,
 RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
 RA Messenguy F., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
 RA Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
 RA Pereira J., Philippsen P., Pierard A., Planta R.J., Plevani P.,
 RA Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
 RA Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
 RA Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
 RA Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C., Thierry A.,
 RA Spiegelberg R., Scateva L.I., Steensma H.F., Steiner S., Vetter I.,
 RA Thiegs G., Triano L.N., Urrestazu L.A., Valle G., Vetter I.,
 RA van Vliet-Reedijk J.C., Volckaert G., Vreken P., Warrington J.R.,
 RA van Wietsteld D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
 RA Zimmermann F.K., Sgourou J.G.;
 RT "The complete DNA sequence of yeast chromosome III.";
 RL Nature 357:38-46(1992).
 RN [3]
 RP REVISIONS.
 RA Gromadka R.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP REVISIONS.
 RA Valles G., Volckaerts G.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Co-assembles with BUD4 at bud sites. BUD4 and BUD3 may
 CC cooperate to recognize a spatial landmark (the neck filaments)
 CC during mitosis and they subsequently become a landmark for
 CC establishing the axial budding pattern in G1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U17580; AAA86315.1;
 CC EMBL; X59720; CAA42346.2;
 CC PIR; S74285; S74285.
 CC PIR; S74286; S74286.
 CC Germonline; 138855;
 CC SGD; S0000520; BUD3.
 CC InterPro; IPR000219; RhoGEF.
 CC SMART; SM00325; RhoGEF; 1.
 CC Cell cycle.

```

SQ SEQUENCE 1636 AA; 184717 MW; 9E4E46BAC3A3F69 CRC64;
Query Match 54.2%; Score 39; DB 1; Length 1636;
Best Local Similarity 58.3%; Pred. No. 2e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PAVVGLSPGQEQ 13
||: |||: ||
Db 112 PATENLSPSQEQ 123

RESULT 14
COBL_MYCTU STANDARD; PRT; 390 AA.
ID COBL_MYCTU
AC Q10671;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Precorrin-6Y C5,15-methyltransferase [decarboxylating] (EC 2.1.1.132)
DE (Precorrin-6 methyltransferase) (Precorrin-6Y methylase).
GN COBL OR RV2072C OR MT2132 OR MTCY49.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
CC -!- FUNCTION: CATALYZES THE METHYLATION OF BOTH C-5 AND C-15 IN
CC PRECORRIN-6Y TO FORM PRECORRIN-8X.
CC -!- CATALYTIC ACTIVITY: 2 S-adenosyl-L-methionine + precorrin-6Y = 2
CC S-adenosyl-L-homocysteine + precorrin-8X + (CO)2.
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: TO S.TYRIMURUM CBIE; ALSO, LOW, TO OTHER
CC METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 273966; CAA38225.1; -.
CC DR EMBL; AE007063; AAK46412.1; -.
CC DR PIR; C70765; C70765.
CC DR TIGR; MT2132; -.
CC TubercuList; RV2072c; -.

DR InterPro; IPR006365; CoBL.
DR InterPro; IPR000878; Cor/pox Metransf.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF00590; TP_methylase; 1.
DR TIGRFAMS; TIGR01468; cobL cbiet; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Complete proteome.
FT CONFLICT 205 L -> P (IN REF. 2).
FT CONFLICT 327 327 D -> H (IN REF. 2).
SQ SEQUENCE 390 AA; 41854 MW; FB42EFF7562F21F3 CRC64;

Query Match 52.8%; Score 38; DB 1; Length 390;
Best Local Similarity 88.9%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPAVVGGLSP 9
|||: |||
Db 55 LPAVQGLSP 63

RESULT 15
RSI_LEULA STANDARD; PRT; 429 AA.
ID RSI_LEULA
AC P50889; P71450;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 40S ribosomal protein S1.
GN RPS1.
OS Leuconostoc lactis.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1246;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97186703; PubMed=9034319;
RA Yamit-Hezi A., Levy Z., Neuman S., Nudel U.;
RA "A Leuconostoc lactis protein with homology to ribosomal protein S1
RT shares common epitopes and common DNA binding properties with a
RT mammalian DNA binding nuclear factor.";
RL Gene 185:99-103(1997).
RN [2]
RP SEQUENCE OF 24-429 FROM N.A.
RX MEDLINE=95237615; PubMed=7721096;
RA Eklund E.A., Lee S.W., Skalniak D.G.;
RA "Cloning of a cDNA encoding a human DNA-binding protein similar to
RT ribosomal protein S1.";
RL Gene 155:231-235(1995).
RN [3]
RP SEQUENCE OF 78-429 FROM N.A.
RX MEDLINE=96164600; PubMed=8568274;
RA Tsuzaka K., Leu A.K., Frank M.B., Movafagh B.F., Koscec M.,
RA Winkler T.H., Kalden J.R., Reichlin M.;
RT "Lupus autoantibodies to double-stranded DNA cross-react with
RT ribosomal protein S1.";
RL J. Immunol. 156:1668-1675(1996).
CC -!- FUNCTION: EXHIBITS PREFERENTIAL BINDING TO SINGLE-STRANDED AND
CC DOUBLE-STRANDED DNA AND A LOW BINDING AFFINITY FOR RNA.
CC -!- SIMILARITY: Belongs to the S1P family of ribosomal proteins.
CC -!- SIMILARITY: Contains 4 S1 motif domains.
CC -!- CAUTION: WAS ORIGINALLY (REF.2 AND REF.3) THOUGHT TO ORIGINATE
CC FROM HUMAN BUT IS MOST PROBABLY THE RESULT OF A CDNA LIBRARY
CC CONTAMINATION BY L.LACTIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U24086; AAB08978.1; -.
CC DR EMBL; U05589; AAA77669.1; -.

```

```

DR EMBL; U27517; AAA97575.1; -.
DR HSP; P05055; LSRO.
DR InterPro; IPR008994; Nucleic acid OB.
DR InterPro; IPR00110; Ribosomal_S1.
DR InterPro; IPR003029; S1.
DR Pfam; PF00575; S1; 4.
DR PRINTS; PR00681; RIBOSOMALS1.
DR SMART; SM00316; S1; 4.
DR PROSITE; PS0126; S1; 4.
KW Ribosomal protein; Repeat; RNA-binding.
FT DOMAIN 55 128 S1 MOTIF 1.
FT DOMAIN 144 211 S1 MOTIF 2.
FT DOMAIN 231 299 S1 MOTIF 3.
FT DOMAIN 316 385 S1 MOTIF 4.
FT CONFLICT 24 24 S -> G (IN REF. 2).
FT CONFLICT 122 122 A -> S (IN REF. 3).
FT CONFLICT 217 217 L -> R (IN REF. 2 AND 3).
SQ SEQUENCE 429 AA; 46386 MW; 92AC82605F39DDFC CRC64;

Query Match 52.8%; Score 38; DB 1; Length 429;
Best Local Similarity 80.0%; Pred.No. 80;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 AVVGLSPGEQ 12
Db |||||
71 AVVGLSTGEE 80

```

Search completed: May 7, 2004, 12:34:33
Job time : 7.04 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 26.18 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-12

Perfect score: 72

Sequence: 1 LPAVVGSLSPGEQY 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_ivirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	68.1	41	11	Q8K408
2	44	61.1	381	16	Q9RL5
3	43.5	60.4	614	16	Q8Y280
4	43	59.7	821	17	Q9HPR8
5	42	58.3	1541	6	Q8HXL3
6	41	56.9	228	17	Q9HKE3
7	41	56.9	326	1	Q9UXP0
8	41	56.9	579	4	Q8N158
9	41	56.9	877	2	Q7X128
10	41	56.9	11096	2	Q9L4W3
11	40	55.6	154	17	Q8ZYJ3
12	40	55.6	227	17	Q97WD1
13	40	55.6	242	16	Q7WCX1
14	40	55.6	261	16	Q8NQA2
15	40	55.6	298	16	Q7W5D6
16	40	55.6	298	16	Q7VTI2

17	40	55.6	353	17	Q8ZU64
18	40	55.6	355	10	Q947A7
19	40	55.6	357	4	Q8NEX1
20	40	55.6	358	4	Q9NWD0
21	40	55.6	366	4	Q9NWM3
22	40	55.6	428	16	Q92AG9
23	40	55.6	428	16	Q8Y661
24	40	55.6	485	10	Q7XNE4
25	40	55.6	540	16	Q9RR71
26	40	55.6	661	10	Q8S8J8
27	40	55.6	665	16	Q8XR57
28	40	55.6	693	2	Q34302
29	40	55.6	753	16	Q89T31
30	39	54.2	148	16	Q8FY0
31	39	54.2	197	17	Q8FWI9
32	39	54.2	209	17	Q8TTT1
33	39	54.2	209	17	Q8TH81
34	39	54.2	277	16	Q9KQK6
35	39	54.2	277	16	Q87RD3
36	39	54.2	284	16	Q7VFI5
37	39	54.2	319	16	Q8ZKQ5
38	39	54.2	319	16	Q8ZKX4
39	39	54.2	407	2	Q9LCW0
40	39	54.2	589	16	Q7UPH0
41	39	54.2	626	16	Q7VWC3
42	39	54.2	627	16	Q7WH05
43	39	54.2	627	16	Q7W9Q3
44	39	54.2	1400	11	Q7TWE1
45	38.5	53.5	94	16	Q88F51

ALIGNMENTS

RESULT 1

Q8K408 PRELIMINARY; PRT; 41 AA.
ID Q8K408
AC Q8K408;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW.tl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E., Sood R., Hashiramoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (CSF1) causes
osteopetrosis in the tl rat.";
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AAM54137.1; -.
SQ SEQUENCE 41 AA; 4178 MW; 1D342C19BD18AA41 CRC64;

Query Match 68.1%; Score 49; DB 11; Length 41;

Best Local Similarity 76.9%; Pred. No. 0.46; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0;

QY 1 LPAVVGSLSPGEQY 13

Db 21 LPAAGLSPEQY 33

RESULT 2

Q9RL5 PRELIMINARY; PRT; 381 AA.

ID Q9RL5

AC Q9RL5; Q7W5D6

DT 01-MAY-2000 (TREMBLrel. 13, Created)


```

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450.
GN DR2473.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heideberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF002076; AAF12016.1; -.
DR PIR; F75270; F75270.
DR TIGR; DR2473; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ SEQUENCE 381 AA; 41940 MW; F191EA69F1797B53 CRC64;

Query Match 61.1%; Score 44; DB 16; Length 381;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAVVGSLSP 9 PRT; 614 AA.
Db 52 LPAVVGSLSP 60

RESULT 3
QY280
ID AC Q8Y280; PRELIMINARY; PRT; 614 AA.
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable ATP-binding transport ABC transporter protein.
GN RSC0456 OR RS04444.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QW11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choigne N., Claude-Renard C., Cunnac S., Denange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigüier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646059; CAD13984.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.

```

```

DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
KW Complete proteome.
SQ SEQUENCE 614 AA; 69240 MW; E293355B85872142 CRC64;

Query Match 60.4%; Score 43.5; DB 16; Length 614;
Best Local Similarity 45.8%; Pred. No. 76;
Matches 11; Conservative 2; Mismatches 0; Indels 11; Gaps 1;

QY 1 LPAVVG-----LSPGEEQ 13
Db 501 LPAVGLLDEVSNWSLRLSPGEEQ 524

RESULT 4
Q9HPR8
ID Q9HPR8 PRELIMINARY; PRT; 821 AA.
AC Q9HPR8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA helicase.
GN HEL OR VNG1501G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AB005064; AAG19799.1; -.
DR PIR; C84304; C84304.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0007242; P:intracellular signalling cascade; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001478; PDZ.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00228; PDZ; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 821 AA; 89848 MW; C454C76B984A5702 CRC64;

Query Match 59.7%; Score 43; DB 17; Length 821;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AVVGLSPGEEQ 12
Db 326 AVVGLSPAEEQ 335

RESULT 5
Q8HXL3
ID Q8HXL3 PRELIMINARY; PRT; 1541 AA.

```

AC Q9HXL3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21845885; PubMed=11856879;
 RA Drogemuller C., Kuiper H., Voss-Nemitz R., Brenig B., Distl O.,
 RA Leeb T.;
 RT "Molecular characterization and chromosome assignment of the porcine
 RT gene COX7A1 coding for the muscle specific cytochrome c oxidase
 RT subunit VIIA-M";
 RL Cytogenet. Cell Genet. 94:190-193 (2001).
 RN [2]
 RN
 RP SEQUENCE FROM N.A.
 RA Leeb T.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ410870; CAD56046.1; -;
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 12.
 DR PROSITE; PS00082; WD_REPEATS_2; 1.
 DR PROSITE; PS02984; WD_REPEATS_REGION; 3.
 KW Hypothetical protein.
 SQ SEQUENCE 1541 AA; 168898 MW; 81B8882854FAF4F1E CRC64;
 Query Match 58.3%; Score 42; DB 6; Length 1541;
 Best Local Similarity 58.3%; Pred. No. 3.7e+02;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LPAVVLSPGGEQ 12
 Db 786 MFSEISLSPGGEQ 797
 : : : |||||

RESULT 6
 Q9HKE3 PRELIMINARY; PRT; 228 AA.
 AC Q9HKE3;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Precorrin-2 methyltransferase related protein.
 GN TA0658.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmatia; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=2303;
 RN [1]
 RN
 RP SEQUENCE FROM N.A.
 RX STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frisman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 RT acidophilum.";
 RL Nature 407:508-513 (2000).
 DR EMBL; AL445065; CAC11796.1; -;
 DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
 DR InterPro; IPR006364; Cobi CbiL.
 DR InterPro; IPR000878; Cor/For Metransf.
 DR InterPro; IPR003043; Uropor. Metransf.
 DR Pfam; PF00590; TP_methylase; 1.
 DR TIGRfam; TIGR01467; cobi cbiL; 1.
 DR PROSITE; PS00839; SUMT 1; 1.
 DR Transferase; Methyltransferase; Complete proteome.

SQ SEQUENCE 228 AA; 25084 MW; 11ABD8B68192A67C CRC64;
 Query Match 56.9%; Score 41; DB 17; Length 228;
 Best Local Similarity 63.6%; Pred. No. 70;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 VVGLSPGGEQ 14
 Db 5 VVGLPGGPDY 15
 : : : |||||

RESULT 7
 Q9UXP0 PRELIMINARY; PRT; 326 AA.
 AC Q9UXP0;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE F420-dependent N5,N10-methylene-tetrahydromethanopterin reductase,
 DE putative.
 GN PFDA.
 OS Methanobolus tindarius.
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanobolus.
 OX NCBI_TaxID=2221;
 RN [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 2278;
 RA Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
 RA Gottschalk G., Blaut M.;
 RT "The F420H2-dehydrogenase from Methanobolus tindarius: Cloning of the
 RT ffd operon and expression of the genes in Escherichia coli.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ011519; CAB56639.1; -;
 DR PIR; T45226; T45226.
 DR InterPro; IPR002103; Bac_luciferase.
 DR Pfam; PF00296; bac_luciferase; 1.
 SQ SEQUENCE 326 AA; 34043 MW; 16F3AB9733A45D82 CRC64;

Query Match 56.9%; Score 41; DB 1; Length 326;
 Best Local Similarity 70.0%; Pred. No. 1e+02;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AVVGLSPGGEQ 12
 Db 88 ALLGLPGGEQ 97
 : : : |||||

RESULT 8
 Q8N158 PRELIMINARY; PRT; 579 AA.
 ID Q8N158;
 AC Q8N158;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Similar to cerebroglycan (Hypothetical protein FLJ38962).
 GN DKFZP547M109.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN
 RP SEQUENCE FROM N.A.
 RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamiyama K.,
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1971;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 11455;
RC MEDLINE=20334850; PubMed=10873841;
RX Brautaset T., Sekurova O.N., Sletta H., Ellingse T.E., Strom A.R.,
RA Vailla S., Zotchev S.B.;
RA "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway.";
RL Chem. Biol. 7:395-403(2000).
DR EMBL; AF263912; RAE71776.1; -.
DR HSSP; P25715; LMLA.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. .; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; AC trans.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; FAbD.
DR InterPro; IPR000794; Ketoacyl_synth.
DR InterPro; IPR006162; Ppantne S.
DR InterPro; IPR006163; Pp bind.
DR Pfam; PF006598; Acyl trans; 6.
DR Pfam; PF001107; ADH_Zinc_N; 1.
DR Pfam; PF001109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fAbD; 6.
DR PROSITE; PS00075; ACP DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 6.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 5.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 11096 AA; 1150415 MW; 776CAEAFCAES51DD CRC64;

Query Match 56.9%; Score 41; DB 2; Length 11096;
Best Local Similarity 63.6%; Pred. No. 4.7e+03;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps

QY 2 PAVVGLSPGEQ 12
DB 4998 PEVTGLAPGDQ 5008

RESULT 11
Q82XJ3 PRELIMINARY; PRT; 154 AA.
ID Q82XJ3
AC Q82XJ3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PAE0746.
GN PAE0746.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OC NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RA "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
RL EMBL; AE009783; RAU63000.1; -.
SQ Hypothetical protein; Complete proteome.
KW SEQUENCE 154 AA; 1717403 MW; 1C3D8BCB40324766 CRC64;

```

Query Match          55.6%; Score 40; DB 17; Length 154;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 10; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY      1 LPVVVGLS--PGEQE 13
      |||: |||: |||:
Db      32 LPDDVGISYTPGEQE 46

RESULT 12
Q97WD1          PRELIMINARY;      PRT;      227 AA.
AC Q97WD1;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Cobalamin biosynthesis precorrin-6B methylase, putative
DE (cblE).
GN CBIE OR SSO2296.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; Pubmed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006833; AAK42457.1; -.
DR PIR; B90400; B90400.
DR GO; GO:0008168; P:methyltransferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000878; Cor/por_Mettransf.
DR Pfam; PF00590; TP_methylase; 1.
KW Methyltransferase; Complete proteome.
SQ SEQUENCE 227 AA; 25548 MW; 0267F09491F2DCED CRC64;

Query Match          55.6%; Score 40; DB 17; Length 227;
Best Local Similarity 54.5%; Pred. No. 1e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      4 VVGLSPGGEQY 14
      |||: |||: |||:
Db      10 IVGVGPDPEY 20

RESULT 13
Q7WCX1          PRELIMINARY;      PRT;      242 AA.
AC Q7WCX1;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Bacteriophage-related DNA polymerase.
GN B3808.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; Pubmed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,

```

```

RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin I., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640448; CAE35782.1; -.
KW Complete proteome.
SQ SEQUENCE 242 AA; 25627 MW; EECAD9B319823284 CRC64;

```

```

Query Match          55.6%; Score 40; DB 16; Length 242;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      4 VVGLSPGGEQ 13
      |||: |||: |||:
Db      96 VVGEAPGGEQ 105

```

```

RESULT 14
Q8NQA2          PRELIMINARY;      PRT;      261 AA.
AC Q8NQA2;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Hypothetical protein Cg11536.
GN CG11536.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AP005278; BAB98929.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 261 AA; 27957 MW; 4D8D51D4DCA3A210 CRC64;

```

```

Query Match          55.6%; Score 40; DB 16; Length 261;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 LPVVVGLSPGGEQ 13
      |||: |||: |||:
Db      137 LPATVSPGEAD 149

```

```

RESULT 15
Q7WSD6          PRELIMINARY;      PRT;      298 AA.
AC Q7WSD6;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Bacteriophage-related DNA polymerase.
GN BPP3357.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;

```

```

RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebatnia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norbertczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR ENBL; BX640433; CAE38642.1; -.
KW Complete proteome.
SQ SEQUENCE 298 AA; 31202 MW; A858C8199E2AC8B7 CRC64;

Query Match 55.6%; Score 40; DB 16; Length 298;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVGLSPGGEQ 13
Db 152 VVGEAPGGEQ 161
||| :|||
||| :|||

```

Search completed: May 7, 2004, 12:37:54
Job time : 28.3467 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 39.06 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-12

Perfect score: 72

Sequence: 1 LPAVGLSPGEQY 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Genesep29Jan04:*

1: Genesep1980s:*

2: Genesep1990s:*

3: Genesep2000s:*

4: Genesep2001s:*

5: Genesep2002s:*

6: Genesep2003as:*

7: Genesep2003bs:*

8: Genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	100.0	14	3	AAy84266 Peptide d
2	72	100.0	20	3	AAy84265 Truncated
3	72	100.0	25	3	AAy84264 Peptide o
4	68	94.4	13	3	AAy84267 Peptide d
5	65	90.3	13	3	AAy84268 Peptide d
6	65	90.3	15	3	AAy84269 Peptide d
7	45	62.5	234	4	AAy84268 Human imm
8	42	58.3	470	6	AAE34724 Streptomy
9	42	58.3	475	6	AAE34732 Streptomy
10	42	58.3	475	6	AAE34729 Streptomy
11	41	56.9	106	5	ABJ10397 Mutant an
12	41	56.9	106	5	ABJ10395 Novel hum
13	41	56.9	306	4	ABG19552 Human NOV
14	41	56.9	530	7	ADD49105 Human NOV
15	41	56.9	549	7	ADD49091 Human NOV
16	41	56.9	579	5	ABG70277 Human Gly
17	41	56.9	579	5	ABG97356 Human GCD
18	41	56.9	579	6	ABR39111 Human GPC
19	41	56.9	579	7	ADD49087 Human NOV
20	41	56.9	579	7	ADD49107 Human NOV
21	41	56.9	579	7	ADD49089 Human NOV
22	41	56.9	592	7	ADD49099 Human NOV
23	41	56.9	11096	4	AAE10129 Streptomy
24	40	55.6	261	4	AAG91441 C glutami
25	40	55.6	287	6	ABU17217 Protein e

26	40	55.6	303	6	ADA36829	Acinetoba
27	40	55.6	306	4	ABG24698	Novel hum
28	40	55.6	358	4	AAy92530	Human pro
29	40	55.6	359	5	ABB97563	Novel hum
30	40	55.6	428	5	ABB47705	Listeria
31	40	55.6	668	3	AAy57055	Human pro
32	39	54.2	18	2	AAy41875	Rheumatoi
33	39	54.2	18	4	AAU25388	Schizophr
34	39	54.2	18	4	AAU15732	Schizophr
35	39	54.2	18	5	ABG78871	Multiple
36	39	54.2	23	5	ABP62618	Human imm
37	39	54.2	74	6	ABJ18694	Antibody
38	39	54.2	88	3	AAy56655	Partial p
39	39	54.2	93	5	AAU80983	Human ant
40	39	54.2	94	7	ADD69248	Human lig
41	39	54.2	95	6	ABO27154	Human ger
42	39	54.2	95	6	ABO27153	Human ger
43	39	54.2	96	6	ABO27155	Human ger
44	39	54.2	96	6	ABO27150	Human ger
45	39	54.2	100	5	AAE23987	Human MOG

ALIGNMENTS

RESULT 1

AAy84266

ID AAY84266 standard; peptide; 14 AA.

XX

AC AAY84266;

XX

DT 12-JUL-2000 (first entry)

XX

DE Peptide derived from macrophage colony stimulating gene alternative ORF.

XX

KW tumour rejection antigen; macrophage colony stimulating gene;

KW macrophage-colony stimulating factor; antigen presenting cell;

KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX

OS Synthetic.

OS Homo sapiens.

XX

FN WO200013699-A1.

XX

PD 16-MAR-2000.

XX

PF 03-SEP-1999; 99WO-US020344.

XX

PR 04-SEP-1998; 98US-0099077P.

XX

PA (LUDW-) LUDWIG INST CANCER RES.

XX

PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX

DR WPI; 2000-256859/22.

XX

PT Isolated polypeptide used to treat subjects having a disorder

PT characterized by expression of alternative open reading frame macrophage-

PT colony stimulating factor comprises 25 amino acid residue sequence.

XX

Claim 2; Page 39; 7app; English.

XX

CC The present sequence represents a peptide which is derived from a tumour

CC rejection antigen precursor encoded by an alternative open reading frame

CC (ORF) of human macrophage colony stimulating gene. Peptides derived from

CC the alternative ORF of macrophage-colony stimulating factor, when

CC presented by an antigen presenting cell having a human leukocyte antigen

CC (HLA) class I molecule, effectively induce the activation and

CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic

CC acids derived from the alternate ORF of macrophage-colony stimulating

CC factor are useful for enriching selectively a population of T lymphocytes

CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder

CC characterized by expression of the polypeptide, and for identifying

KW macrophage-colony stimulating factor; antigen presenting cell;
 XX human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

OS Synthetic.
 OS Homo sapiens.

PN WO200013699-A1.

XX 16-MAR-2000.

PD 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

DR WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 13 AA;

Query Match 94.4%; Score 68; DB 3; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.00026; Indels 0; Gaps 0;
 Matches 13; Conservative 0; Mismatches 0;

OY 2 PAVVGLSPGEQY 14
 DB 1 PAVVGLSPGEQY 13
 |||||

RESULT 5

AA84268
 ID AAY84268 standard; peptide; 13 AA.

XX AC AAY84268;

XX 12-JUL-2000 (first entry)

XX Peptide derived from macrophage colony stimulating gene alternative ORF.
 DE tumour rejection antigen; macrophage colony stimulating gene;
 XX macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Synthetic.
 OS Homo sapiens.

XX WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

PR 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 13 AA;

Query Match 90.3%; Score 65; DB 3; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.0008; Indels 0; Gaps 0;
 Matches 13; Conservative 0; Mismatches 0;

OY 1 LPVVGLSPGEQY 13
 |||||

DB 1 LPVVGLSPGEQY 13
 |||||

RESULT 6

AA84269
 ID AAY84269 standard; peptide; 15 AA.

XX AC AAY84269;

XX 12-JUL-2000 (first entry)

XX Peptide derived from macrophage colony stimulating gene alternative ORF.
 DE tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Synthetic.
 OS Homo sapiens.

XX WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.


```

PS Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics
XX
SQ Sequence 15 AA;

Query Match 90.3%; Score 65; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00094;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAVVGSLSPGEQ 13
   |||||
Db 3 LPAVVGSLSPGEQ 15

RESULT 7
AAB36208
ID AAB36208 standard; protein; 234 AA.
AC AAB36208;
XX
DT 15-FEB-2001 (first entry)
XX
DE Human immune system associated protein HISAP-6.
XX
KW Human; immune system associated protein; HISAP-6; immune disorder;
KW infection; autoimmune disease; cancer.
XX
OS Homo sapiens.
XX
PN US6135941-A.
XX
PD 24-OCT-2000.
XX
PF 27-MAR-1998; 98US-00049672.
XX
PR 27-MAR-1998; 98US-00049672.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;
PI Hillman JL, Au-Young J;
XX
WPI; 2001-030926/04.
DR N-PSDB; AAC66524.
XX
FT New human immune system associated proteins (HISAP) and polynucleotides
FT encoding the HISAP, useful for diagnosing, treating or preventing immune
FT or cell proliferative disorders or infections.
PS Claim 1; Col 59-60; 54pp; English.
XX
CC The present invention provides the coding and protein sequences for a
CC number of human immune system associated proteins (HISAPs). These can be
CC used in the diagnosis and treatment of various autoimmune disorders,
CC infections and cell proliferation diseases. The diseases include AIDS,
CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
CC erythematosus, arteriosclerosis, cirrhosis and cancer
XX
SQ Sequence 234 AA;

Query Match 62.5%; Score 45; DB 4; Length 234;
Best Local Similarity 72.7%; Pred. No. 36;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
   ||:|||||
Db 28 PAIVSLSPGER 38

RESULT 8
AAB34724
ID AAE34724 standard; protein; 470 AA.
XX
AC AAE34724;
XX
DT 14-MAY-2003 (first entry)
XX
DE Streptomyces rimosus ema3 protein.
XX
KW P450 monooxygenase; avermectin; ferredoxin; ferredoxin reductase; enzyme;
KW emamectin; insecticide; ema3 protein.
XX
OS Streptomyces rimosus.
XX
PN WC200292801-A2.
XX
PD 21-NOV-2002.
XX
PF 15-MAY-2002; 2002WO-EP005363.
XX
PR 16-MAY-2001; 2001US-0291149P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
XX
WPI; 2003-140280/13.
DR N-PSDB; AAD53019.
XX
Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
and capable of regioselectively oxidizing alcohol of avermectin useful
for making emamectin from avermectin.
PS Claim 17; Page 107-108; 157pp; English.
XX
CC The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferredoxins and ferredoxin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces rimosus ema3 protein
XX
SQ Sequence 470 AA;

Query Match 58.3%; Score 42; DB 6; Length 470;
Best Local Similarity 72.7%; Pred. No. 2.4e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LPAVVGSLSPGE 11
   ||:|||||
Db 44 LPSIVGLHPGE 54

RESULT 9
AAE34732
ID AAE34732 standard; protein; 475 AA.
XX
AC AAE34732;
XX

```

DT 14-MAY-2003 (first entry)
 XX Streptomyces rimosus ema1 protein.
 DE P450 monooxygenase; avermectin; ferroxidoxin; ferroxidoxin reductase; enzyme;
 KW emamectin; insecticide; ema1 protein.
 XX Streptomyces rimosus.
 OS Streptomyces rimosus.
 XX WO200292801-A2.
 PN 21-NOV-2002.
 XX 15-MAY-2002; 2002WO-EP005363.
 PF 16-MAY-2001; 2001US-0291149P.
 PR (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
 PI Buckel TG;
 XX WPI; 2003-140280/13.
 DR N-PSDB; AAD53027.
 XX Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
 PT and capable of regioselectively oxidizing alcohol of avermectin useful
 PT for making emamectin from avermectin.
 XX Claim 17; Page 121-122; 157pp; English.
 PS The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidising
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferroxidoxins and ferroxidoxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces rimosus ema1 protein
 XX Sequence 475 AA;
 SQ
 Query Match 58.3%; Score 42; DB 6; Length 475;
 Best Local Similarity 72.7%; Pred. No. 2.4e+02;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 LPAVVGLSPGE 11
 ||: ||| |||
 Db 49 LPSYVGLHPGE 59
 RESULT 10
 ID AAE34729 standard; protein; 475 AA.
 XX AAE34729;
 AC AAE34729;
 DT 14-MAY-2003 (first entry)
 XX Streptomyces albofaciens ema8 protein.
 DE P450 monooxygenase; avermectin; ferroxidoxin; ferroxidoxin reductase; enzyme;
 KW emamectin; insecticide; ema8 protein.
 XX Streptomyces albofaciens.
 OS Streptomyces albofaciens.
 XX WO200292801-A2.
 PN 21-NOV-2002.
 PD 15-MAY-2002; 2002WO-EP005363.
 PF 16-MAY-2001; 2001US-0291149P.
 PR

XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
 PI Buckel TG;
 XX WPI; 2003-140280/13.
 DR N-PSDB; AAD53024.
 XX Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
 PT and capable of regioselectively oxidizing alcohol of avermectin useful
 PT for making emamectin from avermectin.
 XX Claim 17; Page 116-117; 157pp; English.
 PS The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidising
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferroxidoxins and ferroxidoxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces albofaciens ema8 protein
 XX Sequence 475 AA;
 SQ
 Query Match 58.3%; Score 42; DB 6; Length 475;
 Best Local Similarity 72.7%; Pred. No. 2.4e+02;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 LPAVVGLSPGE 11
 ||: ||| |||
 Db 49 LPSYVGLHPGE 59
 RESULT 11
 ID ABJ10397 standard; protein; 106 AA.
 XX ABJ10397;
 AC ABJ10397;
 DT 28-NOV-2002 (first entry)
 XX Mutant anti-mesothelin Fv (ST6) variable light chain.
 DE Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.
 KW Unidentified.
 XX Synthetic.
 OS WO200240545-A2.
 PN 23-MAY-2002.
 PD 16-NOV-2001; 2001WO-US043602.
 XX 17-NOV-2000; 2000US-0249805P.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Pastan IH, Onda M, Nagata S, Tsutsumi Y, Vincent JJ, Kreitman RJ;
 PI Vasmatazis G, Lee B;
 XX WPI; 2002-500208/53.
 DR N-PSDB; AHT08085.
 XX Recombinant immunotoxins with reduced liver toxicity for killing
 PT malignant cells bearing antigen comprising mutated framework regions of
 PT antibody heavy and/or light chain or antigen-binding portion of
 PT immunotoxin.
 XX Claim 12; Fig 2; 82pp; English.

CC The invention comprises the amino acid and coding sequences of
 CC recombinant immunotoxin proteins. The immunotoxin proteins of the
 CC invention contain an antibody (or an antigen-binding fragment) with a
 CC substitution of a negatively charged amino acid for an uncharged or
 CC positively charged amino acid. The immunotoxins of the invention have
 CC reduced liver toxicity. The immunotoxins of the invention are useful for
 CC killing a malignant cell (e.g. a cancer cell). The present amino acid
 CC sequence represents a recombinant immunotoxin of the invention
 CC
 XX Sequence 106 AA;

Query Match 56.9%; Score 41; DB 5; Length 106;
 Best Local Similarity 63.6%; Pred. No. 67;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
 Db 8 PAIMSASPGEQ 18
 ||:| |||||

RESULT 12
 ABU10395
 ID ABU10395 standard; protein; 106 AA.

XX AC ABU10395;

XX 28-NOV-2002 (first entry)

DE Mutant anti-Tac Fv (M16) variable light chain.

XX Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.

XX Unidentified.

OS Synthetic.

PN WO200240545-A2.

XX 23-MAY-2002.

XX 16-NOV-2001; 2001WO-US043602.

XX 17-NOV-2000; 2000US-0249805P.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Pastan JH, Onda M, Nagata S, Tsutsumi Y, Vincent JJ, Kreitman RJ;
 PI Vasmatazis G, Lee B;

XX WPI; 2002-500208/53.

DR N-PSDB; APT08083.

XX Recombinant immunotoxins with reduced liver toxicity for killing
 PT malignant cells bearing antigen comprising mutated framework regions of
 PT antibody heavy and/or light chain or antigen-binding portion of
 PT immunotoxin.

XX Claim 8; Fig 2; 82pp; English.

XX The invention comprises the amino acid and coding sequences of
 CC recombinant immunotoxin proteins. The immunotoxin proteins of the
 CC invention contain an antibody (or an antigen-binding fragment) with a
 CC substitution of a negatively charged amino acid for an uncharged or
 CC positively charged amino acid. The immunotoxins of the invention have
 CC reduced liver toxicity. The immunotoxins of the invention are useful for
 CC killing a malignant cell (e.g. a cancer cell). The present amino acid
 CC sequence represents a recombinant immunotoxin of the invention
 CC

XX Sequence 106 AA;

Query Match 56.9%; Score 41; DB 5; Length 106;
 Best Local Similarity 63.6%; Pred. No. 67;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PAVVGLSPGEQ 12
 Db 8 PAIMSASPGEQ 18
 ||:| |||||

RESULT 13

ABG19552

ID ABG19552 standard; protein; 306 AA.

XX AC ABG19552;

XX 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #19543.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS83739.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID NO 49911; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 306 AA;

Query Match 56.9%; Score 41; DB 4; Length 306;
 Best Local Similarity 57.1%; Pred. No. 2.2e+02;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 LPVVGLSPGEQ 14
 ||:| |||||

CC epilepsy, immune disorders (osteoarthritis), hematopoietic disorders,
 CC inflammatory skin disorders, asthma and various dyslipidemias. The coding
 CC sequences and proteins may also be used as targets for the identification
 CC of small molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, in gene therapy, in generation of antibodies that bind
 CC immunospecifically to NOV substances for use in therapeutic or diagnostic
 CC methods.
 CC
 SQ Sequence 530 AA;

Query Match 56.9%; Score 41; DB 7; Length 530;
 Best Local Similarity 75.0%; Pred. No. 4e+02;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LPVVVGLSPGQ 12
 |||||
 Db 416 LPVVVGGSPAEQ 427

RESULT 15
 ADD49091

ID ADD49091 standard; protein; 549 AA.

XX AC ADD49091;

XX 15-JAN-2004 (first entry)

XX Human NOV15d SEQ ID 64.

XX Antidiabetic; anorectic; cardiast; hypotensive; antiarteriosclerotic;
 KW virucide; antibacterial; fungicide; protozoacide; nootropic;
 KW neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;
 KW antiallergic; antiinflammatory; dermatological; antiasthmatic;
 KW antileptic; gene therapy; NOV protein; metabolic disorder; diabetes;
 KW obesity; viral infection; bacterial infection; fungal infection;
 KW helminthic infection; protozoal infection; anorexia; cancer;
 KW cardiovascular disease; hypertension; atherosclerosis;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;
 KW inflammatory skin disorder; asthma; dyslipidemia; human.

OS Homo sapiens.

XX WO2003060149-A2.

XX 24-JUL-2003.

XX 06-JAN-2003; 2003WO-US000252.

XX 04-JAN-2002; 2002US-0345222P.

XX 14-JAN-2002; 2002US-0348693P.

XX 16-JAN-2002; 2002US-0349182P.

XX 17-JAN-2002; 2002US-0349733P.

XX 18-JAN-2002; 2002US-0350263P.

XX 24-JAN-2002; 2002US-0351977P.

XX 28-MAY-2002; 2002US-0383758P.

XX 05-JUN-2002; 2002US-0385969P.

XX 11-JUN-2002; 2002US-0387834P.

XX 17-JUL-2002; 2002US-0396407P.

XX 30-SEP-2002; 2002US-0415115P.

XX 03-JAN-2003; 2003US-00336603.

XX (CURA-) CURAGEN CORP.

XX Grosse WM, Alsobrook JP, Anderson DW, Burgess CE, Edinger SR;
 XX Ellerman K, Furtak K, Gangolli EA, Gerlach VL, Gilbert JA;
 XX Gunther E, Gorman L, Guo X, Ji W, Li L, Miller CE, Padigar M;
 XX Patturajan M, Rastelli L, Macdougall JR, Mishra VS, Smithson G;
 XX Spytek KA, Stone DU, Shenoy SG, Raupier RJ, Vernet CAM, Zhong M;
 XX Malyankar UM, Millet I, Kekuda R;
 XX WPI; 2003-587288/55.

Db 137 LPQAPGLSPGQSW 150

RESULT 14

ADD49105

ID ADD49105 standard; protein; 530 AA.

XX AC ADD49105;

XX 15-JAN-2004 (first entry)

XX Human NOV15k SEQ ID 78.

XX Antidiabetic; anorectic; cardiast; hypotensive; antiarteriosclerotic;
 KW virucide; antibacterial; fungicide; protozoacide; nootropic;
 KW neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;
 KW antiallergic; antiinflammatory; dermatological; antiasthmatic;
 KW antileptic; gene therapy; NOV protein; metabolic disorder; diabetes;
 KW obesity; viral infection; bacterial infection; fungal infection;
 KW helminthic infection; protozoal infection; anorexia; cancer;
 KW cardiovascular disease; hypertension; atherosclerosis;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;
 KW inflammatory skin disorder; asthma; dyslipidemia; human.

XX Homo sapiens.

XX WO2003060149-A2.

XX 24-JUL-2003.

XX 06-JAN-2003; 2003WO-US000252.

XX 04-JAN-2002; 2002US-0345222P.

XX 14-JAN-2002; 2002US-0348693P.

XX 16-JAN-2002; 2002US-0349182P.

XX 17-JAN-2002; 2002US-0349733P.

XX 18-JAN-2002; 2002US-0350263P.

XX 24-JAN-2002; 2002US-0351977P.

XX 28-MAY-2002; 2002US-0383758P.

XX 05-JUN-2002; 2002US-0385969P.

XX 11-JUN-2002; 2002US-0387834P.

XX 17-JUL-2002; 2002US-0396407P.

XX 30-SEP-2002; 2002US-0415115P.

XX 03-JAN-2003; 2003US-00336603.

XX (CURA-) CURAGEN CORP.

XX Grosse WM, Alsobrook JP, Anderson DW, Burgess CE, Edinger SR;
 XX Ellerman K, Furtak K, Gangolli EA, Gerlach VL, Gilbert JA;
 XX Gunther E, Gorman L, Guo X, Ji W, Li L, Miller CE, Padigar M;
 XX Patturajan M, Rastelli L, Macdougall JR, Mishra VS, Smithson G;
 XX Spytek KA, Stone DU, Shenoy SG, Raupier RJ, Vernet CAM, Zhong M;
 XX Malyankar UM, Millet I, Kekuda R;
 XX WPI; 2003-587288/55.

XX N-ESDB; ADD49104.

XX New isolated NOVX polypeptides and polynucleotides, useful for
 XX preventing, diagnosing or treating NOVX-associated disorders, e.g.
 XX osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
 XX asthma, or infections.

XX Claim 1; Page 181-182; 31pp; English.

XX The present invention relates to novel NOV proteins and their coding
 XX sequences (ADD49028-ADD49131). The proteins and coding sequences are
 XX useful in the manufacture of a medicament for treating a syndrome
 XX associated with a human disease, preferably a NOV-associated disorder
 XX such as metabolic disorders, diabetes, obesity, infectious diseases
 XX (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer,
 XX cardiovascular diseases (hypertension, atherosclerosis),
 XX neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,

DR N-PSDB; ADD49090.
XX
PT New isolated NOVX polypeptides and polynucleotides, useful for
PT preventing, diagnosing or treating NOVX-associated disorders, e.g.
PT osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
PT asthma, or infections.
XX
PS Claim 1; Page 177; 311pp; English.
XX
CC The present invention relates to novel NOV proteins and their coding
CC sequences (ADD49028-ADD49131). The proteins and coding sequences are
CC useful in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, preferably a NOV-associated disorder
CC such as metabolic disorders, diabetes, obesity, infectious diseases
CC (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer,
CC cardiovascular diseases (hypertension, atherosclerosis),
CC neurodegenerative disorders (Alzheimer's disease, Parkinson's disease,
CC epilepsy, immune disorders (osteoarthritis), hematopoietic disorders,
CC inflammatory skin disorders, asthma and various dyslipidemias. The coding
CC sequences and proteins may also be used as targets for the identification
CC of small molecules that modulate or inhibit e.g. neurogenesis, cell
CC differentiation, cell proliferation, hematopoiesis, wound healing and
CC angiogenesis, in gene therapy, in generation of antibodies that bind
CC immunospecifically to NOV substances for use in therapeutic or diagnostic
CC methods.
XX
SQ Sequence 549 AA;

Query Match 56.9%; Score 41; DB 7; Length 549;
Best Local Similarity 75.0%; Pred. No. 4.1e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LPAVVGLSPGRQ 12
Db 419 LPPVVGSPAEQ 430

Search completed: May 7, 2004, 12:33:45
Job time : 41.06 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 24.31 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-13
Perfect score: 68
Sequence: 1 PAVVGLSPGREQEY 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	66.2	41	11 Q8K408	Q8K408 rattus norv
2	43	63.2	821	17 Q9HPR8	Q9hpr8 halobacteri
3	41	60.3	228	17 Q9HKE3	Q9hke3 thermoplasm
4	41	60.3	326	1 Q9UXP0	Q9uxp0 methanolobu
5	41	60.3	877	2 Q7X128	Q7xi28 xanthomonas
6	41	60.3	11096	2 Q9L4W3	Q9l4w3 streptomyce
7	40	58.8	227	17 Q97WD1	Q97wd1 sulfolobus
8	40	58.8	242	16 Q7WCX1	Q7wxc1 bordetella
9	40	58.8	238	16 Q7W5D6	Q7w5d6 bordetella
10	40	58.8	298	16 Q7VTI2	Q7vti2 bordetella
11	40	58.8	355	10 Q947A7	Q947a7 nitellopsis
12	40	58.8	357	4 Q8NHX1	Q8nhx1 homo sapien
13	40	58.8	358	4 Q9NWD0	Q9nwd0 homo sapien
14	40	58.8	381	16 Q9RRL5	Q9rrl5 deinococcus
15	40	58.8	386	4 Q9NMW3	Q9nmw3 homo sapien
16	40	58.8	540	16 Q9RR71	Q9rr71 deinococcus

17	40	58.8	661	10 Q8S8J8	Q8s8j8 arabidopsis
18	40	58.8	685	16 Q8XR57	Q8xr57 ralstonia s
19	40	58.8	683	2 Q34302	Q34302 rhizobium e
20	40	58.8	1541	6 Q8HXL3	Q8hxl3 sus scrofa
21	39.5	58.1	614	16 Q8Y280	Q8y280 ralstonia s
22	39	57.4	148	16 Q8FY0	Q8fy0 leptospira
23	39	57.4	197	17 Q8PW19	Q8pw19 methanosarc
24	39	57.4	209	17 Q8TTT1	Q8ttt1 methanosarc
25	39	57.4	209	17 Q8TH81	Q8th81 methanosarc
26	39	57.4	407	2 Q9LCW0	Q9lcw0 streptomyce
27	39	57.4	428	16 Q92AG9	Q92ag9 listeria in
28	39	57.4	428	16 Q8Y661	Q8y661 listeria mo
29	39	57.4	589	16 Q7UPH0	Q7uph0 rhodopirell
30	39	57.4	626	16 Q7VWC3	Q7vwc3 bordetella
31	39	57.4	627	16 Q7WH05	Q7wh05 bordetella
32	39	57.4	627	16 Q7W5Q3	Q7w5q3 bordetella
33	38.5	56.6	94	16 Q88F51	Q88f51 pseudomonas
34	38	55.9	136	11 Q61061	Q61061 mus musculu
35	38	55.9	152	17 Q9V2S2	Q9v2s2 pyrococcus
36	38	55.9	152	17 Q8TZK1	Q8tzk1 pyrococcus
37	38	55.9	155	17 Q57778	Q57778 pyrococcus
38	38	55.9	200	16 Q97LH7	Q97lh7 clostridium
39	38	55.9	203	16 Q8XLI7	Q8xli7 clostridium
40	38	55.9	222	17 Q97WC8	Q97wc8 sulfolobus
41	38	55.9	313	5 Q7YTA2	Q7yta2 glomeris ma
42	38	55.9	325	16 Q9CBS2	Q9cbs2 mycobacteri
43	38	55.9	352	16 Q8G7R2	Q8g7r2 bifidobacte
44	38	55.9	353	17 Q8ZU64	Q8zu64 pyrobaculum
45	38	55.9	404	16 Q897K0	Q897k0 clostridium

ALIGNMENTS

RESULT 1

Q8K408 PRELIMINARY; PRT; 41 AA.
ID Q8K408;
AC Q8K408; (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW tl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.R.; Sood R.; Hashimoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (Csfl) causes
RT osteopetrosis in the tl rat."
RL Biochem Biophys. Res. Commun. 294:1114-1120 (2002).
DR EMBL; AF514357; AAM54137.1;
SQ SEQUENCE 41 AA; 4178 MW; ID342C19BD18AA41 CRC64;

Query Match 66.2%; Score 45; DB 11; Length 41;
Best Local Similarity 75.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAVVGLSPGREQE 12
|||
Db 22 PAAAGLSPREQE 33

RESULT 2

Q9HPR8 PRELIMINARY; PRT; 821 AA.
ID Q9HPR8;
AC Q9HPR8;
DT 01-MAR-2001 (TRENBLrel. 16, Created)

```

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE DNA helicase.
GN HEL OR VNG1501G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlshroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005064; AAC19799.1; -.
DR PIR; C84304; C84304.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001478; PDZ.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00228; PDZ; 1.
DR Helicase; Complete proteome.
KW Helicase; Complete proteome.
SQ SEQUENCE 821 AA; 89848 MW; C454C76B984A5702 CRC64;

Query Match 63.2%; Score 43; DB 17; Length 821;
Best Local Similarity 90.0%; Pred. No. 91;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AVVGLSPGEQ 11
DB 326 AVVGLSPAEG 335

RESULT 3
ID Q9HKE3 PRELIMINARY; PRT; 228 AA.
AC Q9HKE3;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Precorrin-2 methyltransferase related protein.
GN TA0658.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL; AL445085; CAC11796.1; -.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR006364; Cobi_CblI.
DR InterPro; IPR008078; Cor/por Metransf.
DR InterPro; IPR003043; Uropor_Mettransf.
DR Pfam; PF00590; TP_methylase; 1.

TIGRFAMs; TIGR01467; cobi_cblI; 1.
DR PROSITE; PS00839; SUMT_1; 1.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 228 AA; 25084 MW; 11ABD8B68192A67C CRC64;

Query Match 60.3%; Score 41; DB 17; Length 228;
Best Local Similarity 63.6%; Pred. No. 51;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVGLSPGEQEQ 13
DB 5 VVGLSPGDPDY 15

RESULT 4
ID Q9UXP0 PRELIMINARY; PRT; 326 AA.
AC Q9UXP0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE F420-dependent N5,N10-methylene-tetrahydromethanopterin reductase,
DE putative.
GN PFDA.
OS Methanobolus tindarius.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanobolus.
OX NCBI_TaxID=2221;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2278;
RA Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA Gottschalk G., Blaut M.;
RT "The F420H2-dehydrogenase from Methanobolus tindarius: Cloning of the
RT ffd operon and expression of the genes in Escherichia coli."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011519; CAB56639.1; -.
DR PIR; T45226; T45226.
DR InterPro; IPR002103; Bac_luciferase.
DR Pfam; PF00296; bac_luciferase; 1.
SQ SEQUENCE 326 AA; 34043 MW; 16F3AB9733A45D82 CRC64;

Query Match 60.3%; Score 41; DB 1; Length 326;
Best Local Similarity 70.0%; Pred. No. 75;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 AVVGLSPGEQ 11
DB 88 AILGLPGGEQ 97

RESULT 5
ID Q7X128 PRELIMINARY; PRT; 877 AA.
AC Q7X128;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE VirB4.
GN VIRB4.
OS Xanthomonas campestris (pv. citri).
OG Plasmid pXcB.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=346;
RN [1]
RP SEQUENCE FROM N.A.
RA Yuan Q., Brunings A.M., El-Yacoubi B., Shanker S., Gabriel D.W.;
RT "A self-mobilizing plasmid from a South American citrus canker strain
RT carries required pathogenicity gene pthB."
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY228335; AAC72100.1; -.
KW Plasmid.

```

```

SQ SEQUENCE 877 AA; 100012 MW; 83B94EDB23A0728D CRC64;
Query Match 60.3%; Score 41; DB 2; Length 877;
Best Local Similarity 63.6%; Pred. No. 2.2e+02;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVGLSPGQEQY 13
Db 785 VLGLTPGQYEF 795

RESULT 6
Q9L4W3 PRELIMINARY; PRT; 11096 AA.
AC Q9L4W3
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE NYSC.
GN Streptomyces noursei.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1971;
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
RA Brattaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA Valla S., Zetchev S.B.;
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway.";
RL Chem. Biol. 7:395-403(2000).
DR EMBL; AF263912; AAF71776.1; -.
DR HSSP; P25715; IMLA.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0004314; F:[acyl-carrier protein] S-malonyltransferase. .; IEA.
DR GO; GO:0006633; F:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; F:metabolism; IEA.
DR InterPro; IPR001227; AC trans.
DR InterPro; IPR002085; Adh zn family.
DR InterPro; IPR004410; FAb5.
DR InterPro; IPR000794; Ketoacyl synth.
DR InterPro; IPR006162; Ppantne S.
DR InterPro; IPR006163; Pp_bind.
DR Pfam; PF00698; Acyl trans; 6.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRfam; TIGR00128; fadB; 6.
DR PROSITE; PS50075; ACP DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 6.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; 5.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 11096 AA; 1150415 MW; 776CABAFCAE551DD CRC64;

Query Match 60.3%; Score 41; DB 2; Length 11096;
Best Local Similarity 63.6%; Pred. No. 3.4e+03;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVVGLSPGQEQ 11
Db 4998 PEVTLGAPGDQ 5008

RESULT 7
Q97WD1 PRELIMINARY; PRT; 227 AA.
ID Q97WD1
AC Q97WD1

```



```

SQ SEQUENCE 242 AA; 25627 MW; EECAD9B319823284 CRC64;
  Query Match 58.8%; Score 40; DB 16; Length 242;
  Best Local Similarity 80.0%; Pred. No. 82;
  Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVGLSPGQEQ 12
Db 96 VVGEAPGQEQ 105

RESULT 9
Q7W5D6 PRELIMINARY; PRT; 298 AA.
AC Q7W5D6;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Bacteriophage-related DNA polymerase.
GN BPP3357.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Cronin A., Davis P., Doggett J.,
RA Chillingworth T., Collins M., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640433; CAE38642.1; -.
KW Complete proteome.
SQ SEQUENCE 298 AA; 31202 MW; A85C8199E2AC8B7 CRC64;

  Query Match 58.8%; Score 40; DB 16; Length 298;
  Best Local Similarity 80.0%; Pred. No. 1e+02;
  Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVGLSPGQEQ 12
Db 152 VVGEAPGQEQ 161

RESULT 10
Q7VTI2 PRELIMINARY; PRT; 298 AA.
AC Q7VTI2;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Bacteriophage-related DNA polymerase.
GN BP3556.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,

```

```

RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Cronin A., Davis P., Doggett J.,
RA Chillingworth T., Collins M., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640421; CAE43815.1; -.
KW Complete proteome.
SQ SEQUENCE 298 AA; 31231 MW; 9E4BC64B419506F3 CRC64;

  Query Match 58.8%; Score 40; DB 16; Length 298;
  Best Local Similarity 80.0%; Pred. No. 1e+02;
  Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVGLSPGQEQ 12
Db 152 VVGEAPGQEQ 161

RESULT 11
Q947A7 PRELIMINARY; PRT; 355 AA.
AC Q947A7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-
DE glycerate hydro-lyase) (Fragment).
GN ENO.
OS Nitellopsis obtusa.
OC Eukaryota; Viridiplantae; Streptophyta; Charales; Characeae;
OC Nitellopsis.
OX NCBI_TaxID=40811;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X-854;
RX MEDLINE=21437986; PubMed=11526220;
RA Keeling P.J., Palmer J.D.;
RT "Lateral transfer at the gene and subgenic levels in the evolution of
RT eukaryotic enolase.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:10745-10750(2001).
CC -!- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE +
CC H(2)O.
CC -!- COPACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
CC THE DIMER (BY SIMILARITY).
CC -!- PATHWAY: GLYCOLYSIS.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
DR EMBL; AF348916; AAL05455.1; -.
DR GO; GO:0000015; C:phosphopyruvate hydratase complex; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0000287; F:magnesium ion binding; IEA.
DR GO; GO:0004634; F:phosphopyruvate hydratase activity; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR000941; Enolase.
DR Pfam; PF01113; enolase; 1.
DR Pfam; PF03952; enolase; 1.
DR PRINTS; PR00148; ENOLASE.
DR ProDom; PD000902; Enolase; 1.
DR PROSITE; PS00164; ENOLASE; 1.
DR KW Glycolysis; Lyase; Magnesium.
FT NON_TER 1
FT NON_TER 355
SQ SEQUENCE 355 AA; 38295 MW; 400DF160087DE450 CRC64;

  Query Match 58.8%; Score 40; DB 10; Length 355;
  Best Local Similarity 54.5%; Pred. No. 1.2e+02;

```

```

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 PAVVGLSPGEQ 11
Db 28 PAVIGMDPADQ 38

RESULT 12
Q8NHX1 PRELIMINARY; PRT; 357 AA.
AC Q8NHX1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Extracellular signal-related kinase 1b.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Abersold D.M., Yung Y., Seger R.;
RT "Properties of human ERK1b."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY033607; AAK52329.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004707; F:MAP kinase activity; IEA.
DR GO; GO:0004574; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR008349; Erk_1_2_MAPK.
DR InterPro; IPR008350; Erk_3_4_MAPK.
DR InterPro; IPR003527; MAP_Kin.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_Ehr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01770; ERK1ERK2MAPK.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS01351; MAPK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase
SQ SEQUENCE 357 AA; 40662 MW; 58C92773983ADA79 CRC64;

Query Match 58.8%; Score 40; DB 4; Length 357;
Best Local Similarity 72.7%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQ 11
Db 344 PAAVGLGAGEQ 354

RESULT 13
Q9NWD0 PRELIMINARY; PRT; 358 AA.
AC Q9NWD0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue=Embryo;

```

```

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Takiguchi S.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000977; BAA91452.1; -.
DR InterPro; IPR003892; CUE.
DR Pfam; PF02845; CUE; 1.
DR SMART; SM00546; CUE; 1.
KW Hypothetical protein.
SQ SEQUENCE 358 AA; 38760 MW; BCE1AA7E95C73BF0 CRC64;

Query Match 58.8%; Score 40; DB 4; Length 358;
Best Local Similarity 58.3%; Pred. No. 1.3e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQ 12
Db 213 PAMAGPGPDQE 224

RESULT 14
Q9RRL5 PRELIMINARY; PRT; 381 AA.
AC Q9RRL5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450.
GN DR2473.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Farnham W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RA "Genome sequence of the radioresistant bacterium Deinococcus
RA radiodurans R1.";
RT Science 286:1571-1577(1999).
RL -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL; AE002076; AAF12016.1; -.
DR PIR; F75270; F75270.
DR TIGR; DR2473; -.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; F:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ SEQUENCE 381 AA; 41940 MW; F191EA69F179B53 CRC64;

Query Match 58.8%; Score 40; DB 16; Length 381;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAVVGLSP 8
Db 53 PAVVGLSP 60

```

RESULT 15
Q9NWM3 PRELIMINARY; PRT; 386 AA.
ID Q9NWM3
AC Q9NWM3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein FLJ20739.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isegai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000746; BAA91357.1; -
DR Genew; HGNC:19843; CL4orf34.
DR InterPro; IPR003892; CUE.
DR Pfam; PF02845; CUE; 1.
DR SMART; SM00546; CUE; 1.
KW Hypothetical protein.
SQ SEQUENCE 386 AA; 42258 MW; F5530FA47C267895 CRC64;
Query Match 58.8%; Score 40; DB 4; Length 386;
Best Local Similarity 58.3%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 PAVVGLSPGQEE 12
Db 213 PAVAGPGPGDQE 224
Search completed: May 7, 2004, 12:37:55
Job time : 25.4767 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 36.27 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-13
Perfect score: 68
Sequence: 1 PAVVGLSPGEQY 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues
Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Genesep 29Jan04:*
1: Genesep1980s:*
2: Genesep1930s:*
3: Genesep2000s:*
4: Genesep2001s:*
5: Genesep2002s:*
6: Genesep2003as:*
7: Genesep2003bs:*
8: Genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	100.0	13	3 AAY84267	Aay84267 Peptide d
2	68	100.0	14	3 AAY84266	Aay84266 Peptide d
3	68	100.0	20	3 AAY84265	Aay84265 Truncated
4	68	100.0	25	3 AAY84264	Aay84264 Peptide o
5	61	89.7	13	3 AAY84268	Aay84268 Peptide d
6	61	89.7	15	3 AAY84269	Aay84269 Peptide d
7	45	66.2	234	4 AAB36208	Aab36208 Human imm
8	41	60.3	106	5 ABJ10397	Abj10397 Mutant an
9	41	60.3	106	5 ABJ10395	Abj10395 Mutant an
10	41	60.3	11096	4 AAE10129	Aae10129 Streptomy
11	40	58.8	358	4 AAB92530	Aab92530 Human pro
12	40	58.8	359	5 ABB97563	Abb97563 Novel hum
13	40	58.8	668	3 AAB57055	Aab57055 Human pro
14	39	57.4	18	2 AAY41875	Aay41875 Rheumatoi
15	39	57.4	18	4 AAU25388	Aau25388 Schizophr
16	39	57.4	18	4 AAU15732	Aau15732 Schizophr
17	39	57.4	18	5 ABG78871	Abg78871 Multiple
18	39	57.4	23	5 ABF62618	Abf62618 Human-imm
19	39	57.4	74	6 ABL18694	Abj18694 Antibody
20	39	57.4	88	3 AAY56655	Aay56655 Partial p
21	39	57.4	93	5 AAU80983	Aau80983 Human ant
22	39	57.4	94	7 ADD69248	Add69248 Human lig
23	39	57.4	95	6 ABO27154	Abo27154 Human ger
24	39	57.4	95	6 ABO27153	Abo27153 Human ger
25	39	57.4	96	6 ABO27155	Abo27155 Human ger

26	39	57.4	96	6 ABO27150	Abo27150 Human ger
27	39	57.4	100	5 AAE23987	Aae23987 Human MOC
28	39	57.4	104	2 AAW31123	Aaw31123 Alpha lig
29	39	57.4	104	2 AAW26795	Aaw26795 Anti-gp54
30	39	57.4	105	7 ABR61567	Abr61567 HIV-1 neu
31	39	57.4	106	2 AAR41234	Aar41234 Monoclonon
32	39	57.4	106	2 AAW71241	Aaw71241 Light cha
33	39	57.4	106	2 AAE34874	Aae34874 BIWA4 ant
34	39	57.4	106	6 AAE34875	Aae34875 BIWA4 ant
35	39	57.4	106	6 AAE33419	Aae33419 Murine KS
36	39	57.4	106	6 AAE33421	Aae33421 Murine KS
37	39	57.4	106	6 AAO30909	Aao30909 hu-KS ant
38	39	57.4	107	2 AAR32129	Aar32129 Anti-IL2R
39	39	57.4	107	2 AAR37610	Aar37610 B-B10 MAb
40	39	57.4	107	2 AAR50190	Aar50190 Light cha
41	39	57.4	107	2 AAW08949	Aaw08949 Kappa lig
42	39	57.4	107	2 AAW08948	Aaw08948 Kappa lig
43	39	57.4	107	2 AAW24514	Aaw24514 Anti-Fact
44	39	57.4	107	2 AAW24515	Aaw24515 Anti-Fact
45	39	57.4	107	2 AAW24513	Aaw24513 Anti-Fact

ALIGNMENTS

RESULT 1
AAY84267
ID AAY84267 standard; peptide; 13 AA.
XX
AC AAY84267;
XX
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN W0200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.

Example 2; Page 40; 74pp; English.
The present sequence represents a peptide which is derived from a tumour rejection antigen precursor encoded by an alternative open reading frame (ORF) of human macrophage colony stimulating gene. Peptides derived from the alternative ORF of macrophage-colony stimulating factor, when presented by an antigen presenting cell having a human leukocyte antigen (HLA) class I molecule, effectively induce the activation and proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF of macrophage-colony stimulating factor are useful for enriching selectively a population of T lymphocytes with CD8+ T lymphocytes. They are also useful for diagnosing a disorder characterized by expression of the polypeptide, and for identifying

CC functional variants and mimetics

XX Sequence 13 AA;

Query Match 100.0%; Score 68; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.00015; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQY 13
| | | | | | | | | |
Db 1 PAVVGLSPGEQY 13

RESULT 2

AY84266
ID AAY84266 standard; peptide; 14 AA.

AC AAY84266;

DT 12-JUL-2000 (first entry)

Peptide derived from macrophage colony stimulating gene alternative ORF.

KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

OS Synthetic.

OS Homo sapiens.

PN WO200013699-A1.

PD 16-MAR-2000.

PF 03-SEP-1999; 99WO-US020344.

PR 04-SEP-1998; 98US-0099077P.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

DR WPI; 2000-256859/22.

PT Isolated polypeptide used to treat subjects having a disorder
characterized by expression of alternative open reading frame macrophage-
colony stimulating factor comprises 25 amino acid residue sequence.

PS Claim 2; Page 39; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
rejection antigen precursor encoded by an alternative open reading frame
(ORF) of human macrophage colony stimulating gene. Peptides derived from
the alternative ORF of macrophage-colony stimulating factor, when
presented by an antigen presenting cell having a human leukocyte antigen
(HLA) class I molecule, effectively induce the activation and
proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
acids derived from the alternate ORF of macrophage-colony stimulating
factor are useful for enriching selectively a population of T lymphocytes
with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
characterized by expression of the polypeptide, and for identifying
functional variants and mimetics

SQ Sequence 14 AA;

Query Match 100.0%; Score 68; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00016; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQY 13
| | | | | | | | | |
Db 2 PAVVGLSPGEQY 14

RESULT 3

AY84265

ID AAY84265 standard; peptide; 20 AA.

AC AAY84265;

DT 12-JUL-2000 (first entry)

Truncated macrophage colony stimulating factor tumour antigen.

KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

OS Homo sapiens.

PN WO200013699-A1.

PD 16-MAR-2000.

PF 03-SEP-1999; 99WO-US020344.

PR 04-SEP-1998; 98US-0099077P.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

DR WPI; 2000-256859/22.

DR N-PSDB; AAZ99675.

PT Isolated polypeptide used to treat subjects having a disorder
characterized by expression of alternative open reading frame macrophage-
colony stimulating factor comprises 25 amino acid residue sequence.

PS Claim 3; Page 64; 74pp; English.

XX The present sequence represents a truncated tumour rejection antigen
precursor, and is encoded by a truncated alternative open reading frame
(ORF) of human macrophage colony stimulating gene. Peptides derived from
the alternative ORF of macrophage-colony stimulating factor, when
presented by an antigen presenting cell having a human leukocyte antigen
(HLA) class I molecule, effectively induce the activation and
proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
acids derived from the alternate ORF of macrophage-colony stimulating
factor are useful for enriching selectively a population of T lymphocytes
with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
characterized by expression of the polypeptide, and for identifying
functional variants and mimetics

SQ Sequence 20 AA;

Query Match 100.0%; Score 68; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00024; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQY 13
| | | | | | | | | |
Db 5 PAVVGLSPGEQY 17

RESULT 4

AY84264

ID AAY84264 standard; peptide; 25 AA.

AC AAY84264;

DT 12-JUL-2000 (first entry)

Peptide of alternate reading frame of macrophage colony stimulating gene.
Renal cell carcinoma; antigen; cytotoxic T lymphocyte;

KW tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Homo sapiens.

XX WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX N-PSDB; AAZ99672.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Claim 1; Page 64; 74pp; English.

XX The present sequence represents a tumour rejection antigen precursor, and
 CC is encoded by an alternative open reading frame (ORF) of human macrophage
 CC colony stimulating gene. Peptides derived from the alternative ORF of
 CC macrophage-colony stimulating factor, when presented by an antigen
 CC presenting cell having a human leukocyte antigen (HLA) class I molecule,
 CC effectively induce the activation and proliferation of CD8+ cytotoxic T
 CC lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF
 CC of macrophage-colony stimulating factor are useful for enriching
 CC selectively a population of T lymphocytes with CD8+ T lymphocytes. They
 CC are also useful for diagnosing a disorder characterized by expression of
 CC the polypeptide, and for identifying functional variants and mimetics

XX Sequence 25 AA;

Query Match 100.0%; Score 68; DB 3; Length 25;

Best Local Similarity 100.0%; Pred. NO. 0.00031;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGGEQY 13

Db 5 PAVVGLSPGGEQY 17

RESULT 5

AAV84268

ID AAV84268 standard; peptide; 13 AA.

XX AAV84268;

XX 12-JUL-2000 (first entry)

Peptide derived from macrophage colony stimulating gene alternative ORF.

XX tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Synthetic.

XX Homo sapiens.

XX WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

PR 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 13 AA;

Query Match 89.7%; Score 61; DB 3; Length 13;

Best Local Similarity 100.0%; Pred. NO. 0.0022;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGGEQY 12

Db 2 PAVVGLSPGGEQY 13

RESULT 6

AAV84269

ID AAV84269 standard; peptide; 15 AA.

XX AAV84269;

XX 12-JUL-2000 (first entry)

Peptide derived from macrophage colony stimulating gene alternative ORF.

XX tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Synthetic.

XX Homo sapiens.

XX WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

PS Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics
 XX
 SQ Sequence 15 AA;

Query Match 89.7%; Score 61; DB 3; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.0026;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQ 12
 DB 4 PAVVGLSPGEQ 15

RESULT 7
 AAB36208
 ID AAB36208 standard; protein; 234 AA.

AC AAB36208;

DT 15-FEB-2001 (first entry)

DE Human immune system associated protein HISAP-6.

KW Human; immune system associated protein; HISAP-6; immune disorder;
 infection; autoimmune disease; cancer.

OS Homo sapiens.

PN US6135941-A.

PD 24-OCT-2000.

PF 27-MAR-1998; 98US-00049672.

PR 27-MAR-1998; 98US-00049672.

PA (INCY-) INCYTE PHARM INC.

PI Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;

PI Hillman JL, Au-Young J;

DR WPI: 2001-030926/04.

DR N-PSDB; AAC66524.

XX New human immune system associated proteins (HISAP) and polynucleotides
 PT encoding the HISAP, useful for diagnosing, treating or preventing immune
 PT or cell proliferative disorders or infections.

PS Claim 1; Col 59-60; 54pp; English.

XX The present invention provides the coding and protein sequences for a
 CC number of human immune system associated proteins (HISAPs). These can be
 CC used in the diagnosis and treatment of various autoimmune disorders,
 CC infections and cell proliferation diseases. The diseases include AIDS,
 CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
 CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
 CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
 CC erythematosus, arteriosclerosis, cirrhosis and cancer
 XX
 SQ Sequence 234 AA;

Query Match 66.2%; Score 45; DB 4; Length 234;
 Best Local Similarity 72.7%; Pred. No. 25;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQ 11
 DB 28 PAIVLSLSPGER 38

RESULT 8

ABJ10397

ID ABJ10397 standard; protein; 106 AA.

AC ABJ10397;

DT 28-NOV-2002 (first entry)

DE Mutant anti-mesothelin Fv (ST6) variable light chain.

KW Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.

OS Unidentified.

OS Synthetic.

PN WO200240545-A2.

PD 23-MAY-2002.

PF 16-NOV-2001; 2001WO-US043602.

PR 17-NOV-2000; 2000US-0249805P.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Pastan IH, Onda M, Nagata S, Tsutsumi Y, Vincent JJ, Kreitman RJ;

PI Vasmatazis G, Lee B;

DR WPI: 2002-500208/53.

DR N-PSDB; APT08085.

XX Recombinant immunotoxins with reduced liver toxicity for killing
 PT malignant cells bearing antigen comprising mutated framework regions of
 PT antibody heavy and/or light chain or antigen-binding portion of
 PT immunotoxin.

PS Claim 12; Fig 2; 82pp; English.

XX The invention comprises the amino acid and coding sequences of
 CC recombinant immunotoxin proteins. The immunotoxin proteins of the
 CC invention contain an antibody (or an antigen-binding fragment) with a
 CC substitution of a negatively charged amino acid for an uncharged or
 CC positively charged amino acid. The immunotoxins of the invention have
 CC reduced liver toxicity. The immunotoxins of the invention are useful for
 CC killing a malignant cell (e.g. a cancer cell). The present amino acid
 CC sequence represents a recombinant immunotoxin of the invention

XX Sequence 106 AA;

Query Match 60.3%; Score 41; DB 5; Length 106;
 Best Local Similarity 63.6%; Pred. No. 49;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVVGLSPGEQ 11
 DB 8 PAIVLSLSPGER 18

RESULT 9

ABJ10395

ID ABJ10395 standard; protein; 106 AA.

AC ABJ10395;

```

XX 28-NOV-2002 (first entry)
XX
XX
DE Mutant anti-Tac Fv (M16) variable light chain.
XX
XX Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.
XX
XX Unidentified.
XX Synthetic.
XX
XX WO200240545-A2.
XX
XX 23-MAY-2002.
XX
XX 16-NOV-2001; 2001WO-US043602.
XX
XX 17-NOV-2000; 2000US-0249805P.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Pastan IH, Onda M, Nagata S, Tsutsumi Y, Vincent JJ, Kreitman RJ;
PI Vasmatazis G, Lee B;
XX
XX WPI; 2002-500208/53.
XX
XX N-PSDB; ABO8083.
XX
XX Recombinant immunotoxins with reduced liver toxicity for killing
XX malignant cells bearing antigen comprising mutated framework regions of
XX antibody heavy and/or light chain or antigen-binding portion of
XX immunotoxin.
XX
XX Claim 8; Fig 2; 82pp; English.
XX
XX The invention comprises the amino acid and coding sequences of
XX recombinant immunotoxin proteins. The immunotoxin proteins of the
XX invention contain an antibody (or an antigen-binding fragment) with a
XX substitution of a negatively charged amino acid for an uncharged or
XX positively charged amino acid. The immunotoxins of the invention have
XX reduced liver toxicity. The immunotoxins of the invention are useful for
XX killing a malignant cell (e.g. a cancer cell). The present amino acid
XX sequence represents a recombinant immunotoxin of the invention
XX
XX Sequence 106 AA;
XX
XX Query Match 60.3%; Score 41; DB 5; Length 106;
XX Best Local Similarity 63.6%; Pred. No. 49;
XX Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
XX
XX Qy 1 PAVVGLSPGEQ 11
XX ||::|||||
XX Db 8 PAIRASPGEQ 18
XX
XX RESULT 10
XX AAEE10129
XX ID AAEE10129 standard; protein; 11096 AA.
XX
XX AC AAEE10129;
XX
XX DT 29-NOV-2001 (first entry)
XX
XX XX Streptomyces noursei nystatin gene, Nysc.
XX
XX Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic; PKS type I.
XX
XX OS Streptomyces noursei.
XX
XX Key Location/Qualifiers
XX FT Domain 35..455
XX /label= KS3 domain
XX /note= "Ketosynthase (KS) domain"
XX
XX FT Domain 546..858

```

```

FT /label= AT3 domain
FT /note= "Acyltransferase (AT) domain"
FT 872..1073
FT /label= DH3 domain
FT /note= "Dehydratase (DH) domain"
FT 1381..1628
FT /label= KR3 domain
FT /note= "Ketoreductase (KR) domain"
FT 1662..1735
FT /label= ACP3 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 1757..2180
FT /label= KS4 domain
FT /note= "Ketosynthase (KS) domain"
FT 2291..2603
FT /label= AT4 domain
FT /note= "Acyltransferase (AT) domain"
FT 2617..2818
FT /label= DH4 domain
FT /note= "Dehydratase (DH) domain"
FT 3124..3371
FT /label= KR4 domain
FT /note= "Ketoreductase (KR) domain"
FT 3407..3480
FT /label= ACP4 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 3501..3924
FT /label= KS5 domain
FT /note= "Ketosynthase (KS) domain"
FT 4032..4346
FT /label= AT5 domain
FT /note= "Acyltransferase (AT) domain"
FT 4360..4561
FT /label= DH5 domain
FT /note= "Dehydratase (DH) domain"
FT 4953..5239
FT /label= ER5 domain
FT /note= "Enoylreductase (ER) domain"
FT 5248..5495
FT /label= KR5 domain
FT /note= "Ketoreductase (KR) domain"
FT 5528..5601
FT /label= ACP5 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 5623..6046
FT /label= KS6 domain
FT /note= "Ketosynthase (KS) domain"
FT 6166..6478
FT /label= AT6 domain
FT /note= "Acyltransferase (AT) domain"
FT 6492..6704
FT /label= DH6 domain
FT /note= "Dehydratase (DH) domain"
FT 7038..7281
FT /label= KR6 domain
FT /note= "Ketoreductase (KR) domain"
FT 7315..7388
FT /label= ACP6 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 7408..7831
FT /label= KS7 domain
FT /note= "Ketosynthase (KS) domain"
FT 7939..8253
FT /label= AT7 domain
FT /note= "Acyltransferase (AT) domain"
FT 8267..8470
FT /label= DH7 domain
FT /note= "Dehydratase (DH) domain"
FT 8812..9086
FT /label= KR7 domain
FT /note= "Ketoreductase (KR) domain"
FT 9120..9193
FT /label= ACP7 domain

```



```

XX DT 27-JUN-2002 (first entry)
XX DE Novel human protein SEQ ID NO: 831.
XX PF Human; antianaemic; vulnary; antiinflammatory; immunomodulator;
XX KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
XX KW neuroprotective; antiparkinsonian; protein therapy; EST;
XX KW expressed sequence tag.
XX OS Homo sapiens.
XX PN WO200222660-A2.
XX PD 21-MAR-2002.
XX PF 10-SEP-2001; 2001WO-US026015.
XX PF 11-SEP-2000; 2000US-00659671.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX DR WPI; 2002-292408/33.
XX DR N-PSDB; AABN32749.
XX PT An isolated polynucleotide for treating diseases associated with its
XX PT encoded polypeptide such as cancer and multiple sclerosis.
XX PS Claim 20; SEQ ID NO 831; 509pp; English.
XX CC The present invention provides the protein and coding sequences of 444
XX CC novel human proteins. These were isolated from expressed sequences tags
XX CC (ESTs). They can be used to stimulate cell growth, to regulate
XX CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX CC e.g. in burn treatment, to regulate the immune system e.g. to treat
XX CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
XX CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
XX CC rheumatoid arthritis, and to treat nervous system disorders e.g.
XX CC Parkinson's disease. The present sequence is a protein of the invention
XX SQ Sequence 359 AA;

Query Match 58.8%; Score 40; DB 5; Length 359;
Best Local Similarity 58.3%; Pred. No. 2.8e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAVVGLSPGQGE 12
Db 213 PAMAGPGGQGE 224

RESULT 13
AAB57055
ID AAB57055 standard; protein; 668 AA.
XX AC AAB57055;
XX DT 13-MAR-2001 (first entry)
XX DE Human prostate cancer antigen protein sequence SEQ ID NO:1633.
XX KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
XX KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
XX KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
XX KW antibacterial; gene therapy; neural; immune; reproductive; renal;
XX KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX KW wound; infectious disease.
XX OS Homo sapiens.

```

```

XX FN WO2000055174-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US005988.
XX PF 12-MAR-1999; 99US-0124270P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX PI Rosen CA, Ruben SM;
XX DR WPI; 2000-587513/55.
XX DR N-PSDB; AAF16258.
XX PT Prostate cancer associated gene sequences, referred to as prostate cancer
XX PT antigens, useful for treatment, prevention, and diagnosis of disorders
XX PT such as prostate cancer.
XX PS Claim 11; Page 2091-2094; 2338pp; English.
XX CC AAF15566 to AAF16505 encode the human prostate cancer associated
XX CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX CC The prostate cancer antigens can have neuroprotective, cytostatic,
XX CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
XX CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
XX CC and can be used in gene therapy. The prostate cancer antigen
XX CC polynucleotides may be used for detection of prostate cancer, chromosome
XX CC identification, as chromosome markers, and for numerous other diagnostic
XX CC or research purposes. The prostate cancer antigens may be used to treat
XX CC disorders such as neural, immune, muscular, reproductive,
XX CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
XX CC AAB57303 represent sequences used in the exemplification of the present
XX CC invention
XX SQ Sequence 668 AA;

Query Match 58.8%; Score 40; DB 3; Length 668;
Best Local Similarity 72.7%; Pred. No. 5.5e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVGLSPGQGEY 13
Db 27 VVGLSPGQGEY 37

RESULT 14
AAY41875
ID AAY41875 standard; peptide; 18 AA.
XX AC AAY41875;
XX DT 09-DEC-1999 (first entry)
XX DE Rheumatoid arthritis diagnostic protein isoform peptide #26.
XX KW Human; rheumatoid arthritis; RA; diagnosis; RPI; RADP; detection;
XX KW rheumatoid arthritis diagnostic feature; ERPI; synovial fluid;
XX KW rheumatoid arthritis diagnostic protein isoform; screening;
XX KW expression reference protein isoform; prognosis.
XX OS Homo sapiens.
XX PN WO9947925-A2.
XX PD 23-SEP-1999.
XX PF 15-MAR-1999; 99WO-GB0000763.
XX PR 13-MAR-1998; 98GB-00005477.

```

XX (OXFO-) OXFORD GLYSCSCIENCES UK LTD.
 XX
 XX Parekh RB, Patel TP, Townsend RR;
 XX
 XX WPI; 1999-571871/48.
 XX
 XX Diagnosis of human rheumatoid arthritis by two-dimensional
 PT electrophoresis.
 XX
 XX Disclosure; Page 18; 157pp; English.
 XX
 CC A method has been developed for the diagnosis of human rheumatoid
 CC arthritis (RA) using two-dimensional electrophoresis to generate a two-
 CC dimensional array of features. The method can be used for screening,
 CC diagnosis and prognosis of RA in a subject or for monitoring the effect
 CC of an anti-RA drug or therapy administered to a subject. The method
 CC comprises: (a) analysing a sample of serum or plasma and optionally
 CC synovial fluid by two-dimensional electrophoresis, to generate a two-
 CC dimensional array of features; (b) identifying at least one chosen
 CC feature whose relative abundance correlates with the presence or absence
 CC of RA; and (c) comparing the abundance of each chosen feature in the
 CC sample with the abundance of that chosen feature in serum or plasma from
 CC one or more persons without RA, where the relative abundance of the
 CC chosen feature or features in the sample indicates the presence or
 CC absence of RA in the subject. The method can also be used in clinical
 CC studies for testing drugs for therapy of RA, for purification of RA-
 CC diagnostic protein isoforms (RPIs), and for production of antibodies to
 CC RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
 CC compounds that promote or inhibit their activity, which are then used as
 CC RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
 CC protocols. AA41844 to AA42100 represent RPI peptides, AA42101 to
 CC AA42103 represent expression reference protein isoform peptides and
 CC AA225066 to AA225068 represent degenerate probes for RPIs, which are all
 CC used in the exemplification of the present invention
 XX
 XX Sequence 18 AA;
 SQ
 Query Match 57.4%; Score 39; DB 2; Length 18;
 Best Local Similarity 63.6%; Pred. No. 15;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PAVVGLSPGEG 11
 |||:|||||:
 Db 8 PATLSLSPGER 18
 Search completed: May 7, 2004, 12:33:46
 Job time : 37.27 secs

PI Herath HMC, Parekh RB, Rohlf C, Terrett JA, Tyson KL;
 XX
 XX WPI; 2001-570624/64.
 XX
 XX New schizophrenia associated protein isoforms and encoding nucleic acid
 PT molecules, useful for treatment, diagnosis and prognosis of schizophrenia
 PT and screening for potential drugs for treatment and new drug targets.
 XX
 XX Disclosure; Page 41; 148pp; English.
 XX
 CC The sequence represents a schizophrenia-associated protein isoform (SPI).
 CC These protein isoforms, e.g. SPI-206, SPI-238 and SPI-240 are detectable
 CC in cerebrospinal fluid, serum or plasma and are useful markers of
 CC schizophrenia. The sequences can be used for treatment and diagnosis of
 CC schizophrenia, screening, prognosis, monitoring the results of therapy,
 CC identifying patients most likely to respond to a particular therapy and
 CC identification of new targets for drug treatment. SPI DNA is useful as a
 CC nucleic acid probe to detect the presence of nucleic acids or SPIs
 XX
 XX Sequence 18 AA;
 SQ
 Query Match 57.4%; Score 39; DB 4; Length 18;
 Best Local Similarity 63.6%; Pred. No. 15;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PAVVGLSPGEG 11
 |||:|||||:
 Db 8 PATLSLSPGER 18
 Search completed: May 7, 2004, 12:33:46
 Job time : 37.27 secs

XX
 XX (OXFO-) OXFORD GLYSCSCIENCES UK LTD.
 XX
 XX Parekh RB, Patel TP, Townsend RR;
 XX
 XX WPI; 1999-571871/48.
 XX
 XX Diagnosis of human rheumatoid arthritis by two-dimensional
 PT electrophoresis.
 XX
 XX Disclosure; Page 18; 157pp; English.
 XX
 CC A method has been developed for the diagnosis of human rheumatoid
 CC arthritis (RA) using two-dimensional electrophoresis to generate a two-
 CC dimensional array of features. The method can be used for screening,
 CC diagnosis and prognosis of RA in a subject or for monitoring the effect
 CC of an anti-RA drug or therapy administered to a subject. The method
 CC comprises: (a) analysing a sample of serum or plasma and optionally
 CC synovial fluid by two-dimensional electrophoresis, to generate a two-
 CC dimensional array of features; (b) identifying at least one chosen
 CC feature whose relative abundance correlates with the presence or absence
 CC of RA; and (c) comparing the abundance of each chosen feature in the
 CC sample with the abundance of that chosen feature in serum or plasma from
 CC one or more persons without RA, where the relative abundance of the
 CC chosen feature or features in the sample indicates the presence or
 CC absence of RA in the subject. The method can also be used in clinical
 CC studies for testing drugs for therapy of RA, for purification of RA-
 CC diagnostic protein isoforms (RPIs), and for production of antibodies to
 CC RPIs. The RA-diagnostic feature (RADF) proteins can be used to identify
 CC compounds that promote or inhibit their activity, which are then used as
 CC RA drugs. Nucleic acid encoding RADFs can be used in gene therapy
 CC protocols. AA41844 to AA42100 represent RPI peptides, AA42101 to
 CC AA42103 represent expression reference protein isoform peptides and
 CC AA225066 to AA225068 represent degenerate probes for RPIs, which are all
 CC used in the exemplification of the present invention
 XX
 XX Sequence 18 AA;
 SQ
 Query Match 57.4%; Score 39; DB 2; Length 18;
 Best Local Similarity 63.6%; Pred. No. 15;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 PAVVGLSPGEG 11
 |||:|||||:
 Db 8 PATLSLSPGER 18
 Search completed: May 7, 2004, 12:33:46
 Job time : 37.27 secs

RESULT 15
 AAU25388
 ID AAU25388 standard; peptide; 18 AA.
 XX
 XX AAU25388;
 XX
 XX 18-DEC-2001 (first entry)
 XX
 DE Schizophrenia-Associated Protein Isoform (SPI) peptide #617.
 XX
 XX Schizophrenia-associated protein isoform; SPI; SPI-206; SPI-238; SPI-240;
 KW neuroleptic; gene therapy; cerebrospinal fluid; serum; plasma.
 XX
 OS Homo sapiens.
 XX
 XX WO200162785-A2.
 PN
 XX
 XX 30-AUG-2001.
 PD
 XX
 XX 23-FEB-2001; 2001WO-GB000792.
 PF
 XX
 XX 24-FEB-2000; 2000GB-00004415.
 PR
 XX
 XX 28-DEC-2000; 2000US-00750395.
 XX
 XX (OXFO-) OXFORD GLYSCSCIENCES UK LTD.
 PA
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 24.31 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-14

Perfect score: 65

Sequence: 1 LPAVVGSLSPGEQ 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rviro.*
16: sp_bacteriap.*
17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	75.4	41	11 Q8K408	Q8K408 rattus norv
2	44	67.7	381	16 Q9RRL5	Q9RRL5 deinococcus
3	43.5	66.9	614	16 Q8Y280	Q8Y280 raiostonia s
4	43	66.2	821	17 Q9HPR8	Q9HPR8 halobacteri
5	42	64.6	1541	6 Q8HX13	Q8HX13 sus scrofa
6	41	63.1	326	1 Q9UXP0	Q9UXP0 methanolobu
7	41	63.1	579	4 Q8N158	Q8N158 homo sapien
8	41	63.1	11096	2 Q9L4W3	Q9L4W3 streptomyce
9	40	61.5	154	17 Q8ZYJ3	Q8ZYJ3 pyrobaculum
10	40	61.5	242	16 Q7WCX1	Q7WCX1 bordetella
11	40	61.5	261	16 Q8NOA2	Q8NOA2 corynebacte
12	40	61.5	298	16 Q7WSD6	Q7WSD6 bordetella
13	40	61.5	355	10 Q947A7	Q947A7 niteilopsis
14	40	61.5	357	4 Q8NHX1	Q8NHX1 homo sapien
15	40	61.5	357	4 Q8NHX1	Q8NHX1 homo sapien
16	40	61.5	358	4 Q9NWD0	Q9NWD0 homo sapien

17	40	61.5	386	4 Q9NWM3	Q9NWM3 homo sapien
18	40	61.5	428	16 Q9ZAG9	Q9ZAG9 listeria in
19	40	61.5	428	16 Q8Y661	Q8Y661 listeria mo
20	40	61.5	540	16 Q9RR71	Q9RR71 deinococcus
21	40	61.5	753	16 Q89T31	Q89T31 bradyrhizob
22	39	60.0	197	17 Q8PW19	Q8PW19 methanosarc
23	39	60.0	209	17 Q8TWT1	Q8TWT1 methanosarc
24	39	60.0	209	17 Q8TH81	Q8TH81 methanosarc
25	39	60.0	319	16 Q8ZKQ5	Q8ZKQ5 salmonella
26	39	60.0	319	16 Q8Z2X4	Q8Z2X4 salmonella
27	39	60.0	407	2 Q9LCW0	Q9LCW0 streptomyce
28	39	60.0	626	16 Q7WVC3	Q7WVC3 bordetella
29	39	60.0	627	16 Q7WH05	Q7WH05 bordetella
30	39	60.0	627	16 Q7W9Q3	Q7W9Q3 bordetella
31	39	60.0	1400	11 Q7TME1	Q7TME1 mus musculu
32	38.5	59.2	656	16 Q9RW14	Q9RW14 deinococcus
33	38	58.5	62	16 Q7VER3	Q7VER3 mycobacteri
34	38	58.5	136	11 Q61061	Q61061 mus musculu
35	38	58.5	152	17 Q9V2S2	Q9V2S2 pyrococcus
36	38	58.5	152	17 Q8TZK1	Q8TZK1 pyrococcus
37	38	58.5	155	17 Q57778	Q57778 pyrococcus
38	38	58.5	177	16 Q8U7R5	Q8U7R5 agrobacteri
39	38	58.5	212	5 Q9NA60	Q9NA60 caenorhabdi
40	38	58.5	222	17 Q97WC8	Q97WC8 sulfobolus
41	38	58.5	313	5 Q7YTA2	Q7YTA2 glomeris ma
42	38	58.5	325	16 Q9CBS2	Q9CBS2 mycobacteri
43	38	58.5	335	16 Q8ZKR7	Q8ZKR7 streptomyce
44	38	58.5	352	16 Q8G7R2	Q8G7R2 bifidobacte
45	38	58.5	383	16 Q7V8U3	Q7V8U3 prochloroco

ALIGNMENTS

RESULT 1

Q8K408 Q8K408 PRELIMINARY; PRT; 41 AA.
AC Q8K408, 2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW cl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E., Sood R., Hashiramoto A., Hansen C.T., Wilder R.L.,
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (CSF1) causes
RT osteopetrosis in the tl rat."
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AN54137.1; -
SQ SEQUENCE 41 AA; 4178 MW; ID342C19BD18AA41 CRC64;

Query Match 75.4%; Score 49; DB 11; Length 41;
Best Local Similarity 76.9%; Pred. No. 0.44;
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LPAVVGSLSPGEQ 13
||| ||| ||| |||
Db 21 LPAAGLSPREQE 33

RESULT 2

Q9RRL5 Q9RRL5 PRELIMINARY; PRT; 381 AA.
ID Q9RRL5
AC Q9RRL5
DT 01-MAY-2000 (TREMBLrel. 13, Created)

```

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Cytochrome P450.
GN DR2473.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RA "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RT Science 286:1571-1577(1999).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AE002076; AAF12016.1; -.
DR PIR; F75270; F75270.
DR TIGR; DR2473; -.
DR GO; GO:0004437; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase; Complete proteome.
SQ SEQUENCE 361 AA; 41940 MW; F191EA69F1797B53 CRC64;

Query Match 67.7%; Score 44; DB 16; Length 381;
Best Local Similarity 100.0%; Pred. No. 31; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 LPAVVGLSP 9
Db 52 LPAVVGLSP 60

RESULT 3
Q8Y280 ID Q8Y280 PRELIMINARY; PRT; 614 AA.
AC Q8Y280;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Probable ATP-binding transport ABC transporter protein.
GN RSC0456 OR R50444.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Attigenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigquier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RT Nature 415:497-502(2002).
RL Nature 415:497-502(2002).
DR EMBL; AL646059; CAD13984.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.

```

```

DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
KW Complete proteome.
SQ SEQUENCE 614 AA; 69240 MW; E293355B585872142 CRC64;

Query Match 66.2%; Score 43.5; DB 16; Length 614;
Best Local Similarity 45.8%; Pred. No. 63;
Matches 11; Conservative 2; Mismatches 0; Indels 11; Gaps 1;

QY 1 LPAVVG-----LSPGQE 13
Db 501 LPAVLGLLDEVSNWLSLSPGQQ 524

RESULT 4
Q8HPR8 ID Q8HPR8 PRELIMINARY; PRT; 821 AA.
AC Q8HPR8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA helicase.
DS HEL OR VNG1501G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Fohlstroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RL EMBL; AE005064; AAG19799.1; -.
DR PIR; C84304; C84304.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001478; PDZ.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00228; PDZ; 1.
DR KW Helicase; Complete proteome.
SQ SEQUENCE 821 AA; 89848 MW; C454C76B984A5702 CRC64;

Query Match 66.2%; Score 43; DB 17; Length 821;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AVVGLSPGQE 12
Db 326 AVVGLSPAEEQ 335

RESULT 5
Q8HXL3 ID Q8HXL3 PRELIMINARY; PRT; 1541 AA.

```

```

AC Q9HXL3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=41845885; PubMed=11856879;
RA Drogemuller C., Kuiper H., Voss-Nemitz R., Brenig B., Distl O.,
RA Leeb T.;
RT "Molecular characterization and chromosome assignment of the porcine
RT gene COX7A1 coding for the muscle specific cytochrome c oxidase
RT subunit VIIa-M.";
RL Cytogenet. Cell Genet. 94:190-193(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Leeb T.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ410870; CAD56046.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 12.
DR PROSITE; PS50082; WD_REPEATS_2; 1.
DR PROSITE; PS50294; WD_REPEATS_REGION; 3.
KW Hypothetical protein.
SQ SEQUENCE 1541 AA; 168898 MW; 81B8882854FAF1E CRC64;

Query Match 64.6%; Score 42; DB 6; Length 1541;
Best Local Similarity 58.3%; Pred. No. 2.9e+02; Indels 0; Gaps 0;
Matches 7; Conservative 3; Mismatches 2;

Qy 1 LPVVGLSPGEQ 12
::: |||||
Db 786 MPSEISLSPGEQ 797

RESULT 6
Q9UXP0
ID Q9UXP0 PRELIMINARY; PRT; 326 AA.
AC Q9UXP0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F420-dependent N5,N10-methylene-tetrahydromethanopterin reductase,
DE putative.
GN FPD4.
OS Methanobolus tindarius.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanobolus.
OX NCBI_TaxID=2221;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=DSM 2278;
RA Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA Gottschalk G., Blaut M.;
RT "The F420H2-dehydrogenase from Methanobolus tindarius: Cloning of the
RT ffd operon and expression of the genes in Escherichia coli.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011519; CAB56639.1; -.
DR PIR; T45226; T45226.
DR InterPro; IPR002103; Bac_luciferase.
DR Pfam; PF00296; bac_luciferase; 1.
SQ SEQUENCE 326 AA; 34043 MW; 16F3AB9733A45D82 CRC64;

Query Match 63.1%; Score 41; DB 1; Length 326;
Best Local Similarity 70.0%; Pred. No. 88; Indels 0; Gaps 0;
Matches 7; Conservative 2; Mismatches 1;

Qy 3 AVVGLSPGEQ 12
::: |||||

```

```

Db 88 AILGLGPGEQ 97

RESULT 7
Q8N158
ID Q8N158 PRELIMINARY; PRT; 579 AA.
AC Q8N158;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to cerebroglycan (Hypothetical protein FLJ38962).
GN DKFZP547M109.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Bloecher H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027972; AAH27972.1; -.
DR EMBL; AK096281; BAC04745.1; -.
DR EMBL; AL834418; CAD39080.1; -.
DR Gene; HGNC:4450; GPC2.
DR GO; GO:0005578; C:extracellular matrix; IEA.
DR InterPro; IPR001863; Glypican.
DR Pfam; PF01153; Glypican; 1.
DR PROSITE; PS01207; GLYPICAN; 1.
KW Hypothetical protein.
SQ SEQUENCE 579 AA; 62829 MW; 1630E3A22BB83DFA CRC64;

Query Match 63.1%; Score 41; DB 4; Length 579;
Best Local Similarity 75.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 3;

Qy 1 LPVVGLSPGEQ 12
||| ||| ||| |||
Db 439 LPPVVGSPGEQ 450

RESULT 8
Q9L4W3
ID Q9L4W3 PRELIMINARY; PRT; 11096 AA.
AC Q9L4W3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NYSC.
GN NYSC.
OS Streptomyces noursei.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1971;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
RA Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA Valla S., Zotchev S.B.;
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway.";
RL Chem. Biol. 7:395-403(2000).
DR EMBL; AF263912; AAF71776.1; -.
DR HSSP; P25715; 1MLA.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0036740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0004314; F:lacyl-carrier protein; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac trans.
DR InterPro; IPR002085; Adf_zn_family.
DR InterPro; IPR004410; FdbD.
DR InterPro; IPR000794; Ketoacyl_synth.
DR InterPro; IPR006162; Ppatne S.
DR InterPro; IPR006163; Pp bind.
DR Pfam; PF00698; Acyl_transf; 6.
DR Pfam; PF00107; ADH_Zinc_N; 1.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fadb; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; 5.
DR Phosphopantetheine; Transferase.
KW SEQUENCE 11096 AA; 1150415 MW; 776CAEAFCAE551DD CRC64;

Query Match 63.1%; Score 41; DB 2; Length 11096;
Best Local Similarity 63.6%; Pred. No. 3.3e+03;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
Db 4998 PEVTGLAPGDQ 5008

RESULT 9
Q8ZYJ3 PRELIMINARY; PRT; 154 AA.
AC Q8ZYJ3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PAE0746.
GN PAE0746.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AF003783; AL63000.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 154 AA; 17403 MW; 1C3D8BCB40324766 CRC64;

Query Match 61.5%; Score 40; DB 17; Length 154;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 10; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

```

```

QY 1 LPAVVGLS--PGEQ 13
Db 32 LPDVGISYTPGEQ 46

RESULT 10
Q7WCX1 PRELIMINARY; PRT; 242 AA.
AC Q7WCX1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Bacteriophage-related DNA polymerase.
GN BB3808.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Farraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Hason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin I., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640448; CAE35782.1; -.
KW Complete proteome.
SQ SEQUENCE 242 AA; 25627 MW; EECAD9B319823284 CRC64;

Query Match 61.5%; Score 40; DB 16; Length 242;
Best Local Similarity 80.0%; Pred. No. 96;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVGLSPGEQ 13
Db 96 VVGEAPGEQ 105

RESULT 11
Q8NQA2 PRELIMINARY; PRT; 261 AA.
AC Q8NQA2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein Cgl1536.
GN CGL1536.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005278; BAB98929.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 261 AA; 27957 MW; 4D8D51D4DCA3A210 CRC64;

Query Match 61.5%; Score 40; DB 16; Length 261;
Best Local Similarity 53.8%; Pred. No. 1e+02;

```

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LPVVGLSPGEQ 13
 Db 137 LPAATVSPGEAD 149

RESULT 12

Q7W5D6 PRELIMINARY; PRT; 298 AA.
 AC Q7W5D6;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Bacteriophage-related DNA polymerase.
 GN BP3357.
 OS Bordetella parapertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=12822 / ATCC BAA-587;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Saunders M., Squares S., Stevens K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; EX640433; CAE38642.1; -.
 KW Complete proteome.
 SQ SEQUENCE 298 AA; 31202 MW; A858C8199E2AC8B7 CRC64;

Query Match 61.5%; Score 40; DB 16; Length 298;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVGLSPGEQ 13
 Db 152 VVGEAPGEQ 161

RESULT 13

Q7VTI2 PRELIMINARY; PRT; 298 AA.
 AC Q7VTI2;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Bacteriophage-related DNA polymerase.
 GN BP3556.
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parkhill J., Sebaihia M., Frescon A., Murphy L.D., Thomson N.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,

RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; EX640421; CAE43815.1; -.
 KW Complete proteome.
 SQ SEQUENCE 298 AA; 31231 MW; 9E4BC84B419506F3 CRC64;

Query Match 61.5%; Score 40; DB 16; Length 298;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVGLSPGEQ 13
 Db 152 VVGEAPGEQ 161

RESULT 14

Q947A7 PRELIMINARY; PRT; 355 AA.
 AC Q947A7;
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase) (2-phospho-D-
 glycerate hydro-lyase) (Fragment).
 GN ENO.
 OS Nitellopsis obtusa.
 OC Eukaryota; Viridiplantae; Streptophyta; Charales; Characeae;
 OC Nitellopsis.
 OX NCBI_TaxID=40811;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X-854;
 RX MEDLINE=21437986; PubMed=11526220;
 RA Keeling P.J., Palmer J.D.;
 RT "Lateral transfer at the gene and subgenic levels in the evolution of
 eukaryotic enolase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:10745-10750(2001).
 CC -1- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE +
 H(2O).
 CC -1- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
 THE DIMER (BY SIMILARITY).
 CC -1- PATHWAY: GLYCOLYSIS.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.
 DR EMBL; AF348916; AAL05455.1; -.
 DR GO; GO:0000015; C:phosphopyruvate hydratase complex; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0000287; F:magnesium ion binding; IEA.
 DR GO; GO:0004634; F:phosphopyruvate hydratase activity; IEA.
 DR GO; GO:0006096; P:glycolysis; IEA.
 DR InterPro; IPR000941; Enolase.
 DR Pfam; PF00113; enolase; 1.
 DR Pfam; PF03952; enolase_N; 1.
 DR PRINTS; PR00148; ENOLASE.
 DR ProDom; PD000902; Enolase; 1.
 DR PROSITE; PS00164; ENOLASE; 1.
 KW Glycolysis; Lyase; Magnesium.
 FT NON TER 1
 FT NON TER 355 355
 SQ SEQUENCE 355 AA; 38295 MW; 400DF160087DE450 CRC64;

Query Match 61.5%; Score 40; DB 10; Length 355;
 Best Local Similarity 54.5%; Pred. No. 1.4e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
 Db 152 PAVVGLSPGEQ 12

Db 28 PAVIGMDPADQ 38

RESULT 15

Q8NHX1 PRELIMINARY; PRT; 357 AA.
AC Q8NHX1
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Extracellular signal-related kinase 1b.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP Abersold D.M., Yung Y., Seger R.;
RA "Properties of human ERK1b."
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY033607; RAK52329.1; -;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004707; F:MAP kinase activity; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR008349; Erk_1_2_MAPK.
DR InterPro; IPR008350; Erk_3_4_MAPK.
DR InterPro; IPR003527; MAP_kin.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01770; ERKLERK2MAPK.
DR PRINTS; PR01771; ERK3ERK4MAPK.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS01351; MAPK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase.
SQ SEQUENCE 357 AA; 40062 MW; 58C92773983ADA79 CRC64;

Query Match 61.5%; Score 40; DB 4; Length 357;
Best Local Similarity 72.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
Db 344 PAAVGLGAGEQ 354

Search completed: May 7, 2004, 12:37:56
Job time : 25.4767 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 36.27 Seconds
(without alignment)

101.272 Million cell updates/sec

Title: US-09-786-214A-14

Perfect score: 65

Sequence: 1 LPAVVGSLSPGEQE 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Genesep29Jan04.*

1: Genesep1980s.*

2: Genesep1990s.*

3: Genesep2000s.*

4: Genesep2001s.*

5: Genesep2002s.*

6: Genesep2003as.*

7: Genesep2003bs.*

8: Genesep2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	13	3	AAy84268 Peptide d
2	65	100.0	14	3	AAy84266 Peptide d
3	65	100.0	15	3	AAy84269 Peptide d
4	65	100.0	20	3	AAy84265 Truncated
5	65	100.0	25	3	AAy84264 Peptide o
6	61	93.8	13	3	AAy84267 Peptide d
7	45	69.2	234	4	AAAB36208 Human imm
8	42	64.6	470	6	AAE34724 Streptomy
9	42	64.6	475	6	AAE34732 Streptomy
10	42	64.6	475	6	AAE34729 Streptomy
11	41	63.1	106	5	ABJ10397 Mutant an
12	41	63.1	106	5	ABJ10395 Mutant an
13	41	63.1	530	7	ADD49105 Human NOV
14	41	63.1	530	7	ADD49091 Human NOV
15	41	63.1	579	5	ABg70277 Human Gly
16	41	63.1	579	5	ABg97356 Human GPD
17	41	63.1	579	6	ABR39111 Human GPC
18	41	63.1	579	7	ADD49087 Human NOV
19	41	63.1	579	7	ADD49107 Human NOV
20	41	63.1	579	7	ADD49089 Human NOV
21	41	63.1	592	7	ADD49099 Human NOV
22	41	63.1	11096	4	AAE10129 Streptomy
23	40	61.5	261	4	AAg91441 C glutami
24	40	61.5	306	4	ABg24698 Novel hum
25	40	61.5	358	4	AAb92530 Human pro

26	40	61.5	359	5	ABB97563 Novel hum
27	40	61.5	428	5	ABB47705 Listeria
28	39	60.0	18	2	AAy41875 Rheumatoi
29	39	60.0	18	4	AAU25388 Schizophr
30	39	60.0	18	4	AAU15732 Schizophr
31	39	60.0	18	5	ABg78871 Multiple
32	39	60.0	23	5	ABP62618 Human imm
33	39	60.0	74	6	ABJ18694 Antibody
34	39	60.0	88	3	AAy56655 Partial p
35	39	60.0	93	5	AAU80983 Human ant
36	39	60.0	94	7	ADD69248 Human lig
37	39	60.0	95	6	ABO27154 Human ger
38	39	60.0	95	6	ABO27153 Human ger
39	39	60.0	96	6	ABO27155 Human ger
40	39	60.0	96	6	ABO27150 Human ger
41	39	60.0	100	5	AAE23987 Human MCG
42	39	60.0	104	2	AAW31723 Alpha lig
43	39	60.0	104	2	AAW26795 Anti-gp54
44	39	60.0	105	7	ABR61567 HIV-1 neu
45	39	60.0	106	2	AAr41234 Monoclonal

ALIGNMENTS

RESULT 1

AAy84268

ID AAY84268 standard; peptide; 13 AA.

XX AC AAY84268;

XX AC

DT 12-JUL-2000 (first entry)

XX XX

DE Peptide derived from macrophage colony stimulating gene alternative ORF.

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

XX KW

CC functional variants and mimetics

XX SQ Sequence 13 AA;

Query Match 100.0%; Score 65; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.00057;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGQE 13

Db 1 LPAVVGLSPGQE 13

RESULT 2

AAy84266

ID AAY84266 standard; peptide; 14 AA.

XX AC AAY84266;

XX DT 12-JUL-2000 (first entry)

Peptide derived from macrophage colony stimulating gene alternative ORF.

XX DE tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO200013699-A1.

XX PD 16-MAR-2000.

XX PF 03-SEP-1999; 99WO-US020344.

XX PR 04-SEP-1998; 98US-0099077P.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX DR WPI; 2000-256859/22.

XX PT Isolated polypeptide used to treat subjects having a disorder
characterized by expression of alternative open reading frame macrophage-
colony stimulating factor comprises 25 amino acid residue sequence.

XX PS Claim 2; Page 39; 74pp; English.

XX CC The present sequence represents a peptide which is derived from a tumour
rejection antigen precursor encoded by an alternative open reading frame
(ORF) of human macrophage colony stimulating gene. Peptides derived from
the alternative ORF of macrophage-colony stimulating factor, when
presented by an antigen presenting cell having a human leukocyte antigen
(HLA) class I molecule, effectively induce the activation and
proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
acids derived from the alternate ORF of macrophage-colony stimulating
factor are useful for enriching selectively a population of T lymphocytes
with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
characterized by expression of the polypeptide, and for identifying
functional variants and mimetics

XX SQ Sequence 14 AA;

Query Match 100.0%; Score 65; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00061;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGQE 13

Db 1 LPAVVGLSPGQE 13

RESULT 3

AAy84269

ID AAY84269 standard; peptide; 15 AA.

XX AC AAY84269;

XX DT 12-JUL-2000 (first entry)

Peptide derived from macrophage colony stimulating gene alternative ORF.

XX DE tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO200013699-A1.

XX PD 16-MAR-2000.

XX PF 03-SEP-1999; 99WO-US020344.

XX PR 04-SEP-1998; 98US-0099077P.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX DR WPI; 2000-256859/22.

XX PT Isolated polypeptide used to treat subjects having a disorder
characterized by expression of alternative open reading frame macrophage-
colony stimulating factor comprises 25 amino acid residue sequence.

XX PS Example 2; Page 40; 74pp; English.

XX CC The present sequence represents a peptide which is derived from a tumour
rejection antigen precursor encoded by an alternative open reading frame
(ORF) of human macrophage colony stimulating gene. Peptides derived from
the alternative ORF of macrophage-colony stimulating factor, when
presented by an antigen presenting cell having a human leukocyte antigen
(HLA) class I molecule, effectively induce the activation and
proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
acids derived from the alternate ORF of macrophage-colony stimulating
factor are useful for enriching selectively a population of T lymphocytes
with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
characterized by expression of the polypeptide, and for identifying
functional variants and mimetics

XX SQ Sequence 15 AA;

Query Match 100.0%; Score 65; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00066;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPAVVGLSPGQE 13

Db 3 LPAVVGLSPGQE 15

RESULT 4

AAy84265

ID AAY84265 standard; peptide; 20 AA.

XX AC AAY84265;

XX DT 12-JUL-2000 (first entry)

Truncated macrophage colony stimulating factor tumour antigen.

XX KW tumour rejection antigen; macrophage colony stimulating gene;

KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Homo sapiens.

PN WO200013699-A1.

PD 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

DR N-PSDB; AAZ99675.

XX Isolated polypeptide used to treat subjects having a disorder

PT characterized by expression of alternative open reading frame macrophage-

PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Claim 3; Page 64; 74pp; English.

XX The present sequence represents a truncated tumour rejection antigen
 CC precursor, and is encoded by a truncated alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 65; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.0009;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAVWGLSPGEQE 13

Db 4 LPAVWGLSPGEQE 16

RESULT 5

AAV84264

ID AAV84264 standard; peptide; 25 AA.

XX AC AAV84264;

XX 12-JUL-2000 (first entry)

XX Peptide of alternate reading frame of macrophage colony stimulating gene.

DE Renal cell carcinoma; antigen; cytotoxic T lymphocyte;

XX tumour rejection antigen; macrophage colony stimulating gene;

KW macrophage-colony stimulating factor; antigen presenting cell;

KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Homo sapiens.

OS WO200013699-A1.

PN 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX

PR 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

DR N-PSDB; AAZ99672.

XX Isolated polypeptide used to treat subjects having a disorder

PT characterized by expression of alternative open reading frame macrophage-

PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Claim 1; Page 64; 74pp; English.

XX The present sequence represents a tumour rejection antigen precursor, and

CC is encoded by an alternative open reading frame (ORF) of human macrophage

CC colony stimulating gene. Peptides derived from the alternative ORF of

CC macrophage-colony stimulating factor, when presented by an antigen

CC presenting cell having a human leukocyte antigen (HLA) class I molecule,

CC effectively induce the activation and proliferation of CD8+ cytotoxic T

CC lymphocytes. Polypeptide and nucleic acids derived from the alternate ORF

CC of macrophage-colony stimulating factor are useful for enriching

CC selectively a population of T lymphocytes with CD8+ T lymphocytes. They

CC are also useful for diagnosing a disorder characterized by expression of

CC the polypeptide, and for identifying functional variants and mimetics

XX SQ Sequence 25 AA;

Query Match 100.0%; Score 65; DB 3; Length 25;

Best Local Similarity 100.0%; Pred. No. 0.0012;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPAVWGLSPGEQE 13

Db 4 LPAVWGLSPGEQE 16

RESULT 6

AAV84267

ID AAV84267 standard; peptide; 13 AA.

XX AC AAV84267;

XX 12-JUL-2000 (first entry)

XX Peptide derived from macrophage colony stimulating gene alternative ORF.

DE tumour rejection antigen; macrophage colony stimulating gene;

XX macrophage-colony stimulating factor; antigen presenting cell;

KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Synthetic.

OS Homo sapiens.

OS WO200013699-A1.

PN 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder

PT characterized by expression of alternative open reading frame macrophage-

PT colony stimulating factor comprises 25 amino acid residue sequence.

XX

PS Example 2; Page 40; 74pp; English.

CC The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 13 AA;

Query Match 93.8%; Score 61; DB 3; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.0026;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQE 13
 |||||
 Db 1 PAVVGLSPGEQE 12

RESULT 7

AAAB36208
 ID AAB36208 standard; protein; 234 AA.

AC AAB36208;

DT 15-FEB-2001 (first entry)

XX Human immune system associated protein HISAP-6.

XX Human; immune system associated protein; HISAP-6; immune disorder;
 KW infection; autoimmune disease; cancer.

XX Homo sapiens.

PN US6135941-A.

XX 24-OCT-2000.

PF 27-MAR-1998; 98US-00049672.

PR 27-MAR-1998; 98US-00049672.

XX (INCY-) INCYTE PHARM INC.

PI Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;

PI Hillman JL, Au-Young J;

XX WPI: 2001-030926/04.

DR N-PSDB; AAC66524.

XX New human immune system associated proteins (HISAP) and polynucleotides
 PT encoding the HISAP, useful for diagnosing, treating or preventing immune
 PT or cell proliferative disorders or infections.

PS Claim 1; Col 59-60; 54pp; English.

XX The present invention provides the coding and protein sequences for a
 CC number of human immune system associated proteins (HISAPs). These can be
 CC used in the diagnosis and treatment of various autoimmune disorders,
 CC infections and cell proliferation diseases. The diseases include AIDS,
 CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
 CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
 CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
 CC erythematosus, arteriosclerosis, cirrhosis and cancer

XX Sequence 234 AA;

Query Match 69.2%; Score 45; DB 4; Length 234;
 Best Local Similarity 72.7%; Pred. No. 26;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQE 12
 |||||
 Db 28 PAIVSLSPGER 38

RESULT 8

AAE34724
 ID AAE34724 standard; protein; 470 AA.

AC AAE34724;

DT 14-MAY-2003 (first entry)

XX Streptomyces rimosus ema3 protein.

XX P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
 KW emamectin; insecticide; ema3 protein.

XX Streptomyces rimosus.

PN WO200292801-A2.

PD 21-NOV-2002.

PF 15-MAY-2002; 2002WO-EP005363.

XX 16-MAY-2001; 2001US-0291149P.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
 PI Buckel TG;

XX WPI: 2003-140280/13.

DR N-PSDB; AAD53019.

XX Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
 PT and capable of regioselectively oxidizing alcohol of avermectin useful
 PT for making emamectin from avermectin.

PS Claim 17; Page 107-108; 157pp; English.

XX The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidising
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferrodoxins and ferrodoxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces rimosus ema3 protein

XX Sequence 470 AA;

Query Match 64.6%; Score 42; DB 6; Length 470;
 Best Local Similarity 72.7%; Pred. No. 1.8e+02;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LPAVVGLSPGE 11
 |||||
 Db 44 LPSYVGLHPGE 54

RESULT 9

AAE34732
 ID AAE34732 standard; protein; 475 AA.

AC AAE34732;

XX

DT 14-MAY-2003 (first entry)
 XX Streptomyces rimosus small protein.
 DE P450 monooxygenase; avermectin; ferrodioxin; ferrodioxin reductase; enzyme;
 KW emamectin; insecticide; small protein.
 XX Streptomyces rimosus.
 XX WO200292801-A2.
 FN 21-NOV-2002.
 XX 15-MAY-2002; 2002WO-EP005363.
 XX 16-MAY-2001; 2001US-0291149P.
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 PA Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
 PI Buckel TG;
 XX WPI; 2003-140280/13.
 DR N-PSDB; AAD53027.
 XX Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
 PT and capable of regioselectively oxidizing alcohol of avermectin useful
 PT for making emamectin from avermectin.
 XX Claim 17; Page 121-122; 157pp; English.
 CC The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidising
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferrodioxins and ferrodioxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces rimosus small protein
 XX Sequence 475 AA;
 SQ
 Query Match 64.6%; Score 42; DB 6; Length 475;
 Best Local Similarity 72.7%; Pred. NO. 1.8e+02;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 1 LPAVVGLSPGE 11
 Db 49 LPSYVGLHPGE 59
 ||: ||| |||
 RESULT 10
 AAEE34729
 ID AAEE34729 standard; protein; 475 AA.
 AC AAEE34729;
 XX 14-MAY-2003 (first entry)
 XX Streptomyces albofaciens ema8 protein.
 DE P450 monooxygenase; avermectin; ferrodioxin; ferrodioxin reductase; enzyme;
 KW emamectin; insecticide; ema8 protein.
 XX Streptomyces albofaciens.
 XX WO200292801-A2.
 FN 21-NOV-2002.
 XX 15-MAY-2002; 2002WO-EP005363.
 XX 16-MAY-2001; 2001US-0291149P.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
 PI Buckel TG;
 XX WPI; 2003-140280/13.
 DR N-PSDB; AAD53024.
 XX Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
 PT and capable of regioselectively oxidizing alcohol of avermectin useful
 PT for making emamectin from avermectin.
 XX Claim 17; Page 116-117; 157pp; English.
 XX The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidising
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferrodioxins and ferrodioxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces albofaciens ema8 protein
 XX Sequence 475 AA;
 SQ
 Query Match 64.6%; Score 42; DB 6; Length 475;
 Best Local Similarity 72.7%; Pred. NO. 1.8e+02;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 1 LPAVVGLSPGE 11
 Db 49 LPSYVGLHPGE 59
 ||: ||| |||
 RESULT 11
 ABUT10397
 ID ABUT10397 standard; protein; 106 AA.
 AC ABUT10397;
 XX 28-NOV-2002 (first entry)
 XX Mutant anti-mesothelin Fv (ST6) variable light chain.
 DE Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.
 XX Unidentified.
 OS Synthetic.
 XX WO200240545-A2.
 XX 23-MAY-2002.
 XX 16-NOV-2001; 2001WO-US043602.
 XX 17-NOV-2000; 2000US-0249805P.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Pastan IH, Onda M, Nagata S, Teutsami Y, Vincent JJ, Kreitman RJ;
 PI Vasmatazis G, Lee B;
 XX WPI; 2002-500208/53.
 DR N-PSDB; AET08085.
 XX Recombinant immunotoxins with reduced liver toxicity for killing
 PT malignant cells bearing antigen comprising mutated framework regions of
 PT antibody heavy and/or light chain or antigen-binding portion of
 PT immunotoxin.
 XX Claim 12; Fig 2; 82pp; English.

CC The invention comprises the amino acid and coding sequences of
 CC recombinant immunotoxin proteins. The immunotoxin proteins of the
 CC invention contain an antibody (or an antigen-binding fragment) with a
 CC substitution of a negatively charged amino acid for an uncharged or
 CC positively charged amino acid. The immunotoxins of the invention have
 CC reduced liver toxicity. The immunotoxins of the invention are useful for
 CC killing a malignant cell (e.g. a cancer cell). The present amino acid
 CC sequence represents a recombinant immunotoxin of the invention
 XX
 SQ Sequence 106 AA;

Query Match 63.1%; Score 41; DB 5; Length 106;
 Best Local Similarity 63.6%; Pred. No. 51;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
 DB 8 PAIMASPGEQ 18
 ||: |||||

RESULT 12
 ABJ10395

ID ABJ10395 standard; protein; 106 AA.

AC ABJ10395;

DT 28-NOV-2002 (first entry)

XX Mutant anti-Tac Fv (M16) variable light chain.

DE Mutant; mutein; recombinant immunotoxin; reduced liver toxicity; cancer.

XX Unidentified.

OS Synthetic.

XX WO200240545-A2.

XX 23-MAY-2002.

PF 16-NOV-2001; 2001WO-US043602.

XX 17-NOV-2000; 2000US-0249805P.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Pastan IH, Onda M, Nagata S, Tsutsumi Y, Vincent JJ, Kreitman RJ;
 PI Vasmatazis G, Lee B;

XX WPI; 2002-500208/53.

DR N-PSDB; ABT08083.

XX Recombinant immunotoxins with reduced liver toxicity for killing
 PT malignant cells bearing antigen comprising mutated framework regions of
 PT antibody heavy and/or light chain or antigen-binding portion of
 PT immunotoxin.

PS Claim 8; Fig 2; 82pp; English.

XX The invention comprises the amino acid and coding sequences of
 CC recombinant immunotoxin proteins. The immunotoxin proteins of the
 CC invention contain an antibody (or an antigen-binding fragment) with a
 CC substitution of a negatively charged amino acid for an uncharged or
 CC positively charged amino acid. The immunotoxins of the invention have
 CC reduced liver toxicity. The immunotoxins of the invention are useful for
 CC killing a malignant cell (e.g. a cancer cell). The present amino acid
 CC sequence represents a recombinant immunotoxin of the invention
 XX
 SQ Sequence 106 AA;

Query Match 63.1%; Score 41; DB 5; Length 106;
 Best Local Similarity 63.6%; Pred. No. 51;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAVVGLSPGEQ 12
 DB 8 PAIMASPGEQ 18
 ||: |||||

RESULT 13

ADD49105

ID ADD49105 standard; protein; 530 AA.

XX AC ADD49105;

XX 15-JAN-2004 (first entry)

XX Human NOV15k SEQ ID 78.

XX Antidiabetic; anorectic; cardiac; hypotensive; antiarteriosclerotic;
 KW viricide; antibacterial; fungicide; protozoacide; nootropic;
 KW neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;
 KW antiarthritic; antiinflammatory; dermatological; antiasthmatic;
 KW antilipemic; gene therapy; NOV protein; metabolic disorder; diabetes;
 KW obesity; viral infection; bacterial infection; fungal infection;
 KW helminthic infection; protozoal infection; anorexia; cancer;
 KW cardiovascular disease; hypertension; atherosclerosis;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;
 KW inflammatory skin disorder; asthma; dyslipidemia; human.

OS Homo sapiens.

XX WO2003060149-A2.

XX 24-JUL-2003.

XX 06-JAN-2003; 2003WO-US000252.

XX 04-JAN-2002; 2002US-0345222P.

PR 14-JAN-2002; 2002US-0348693P.

PR 16-JAN-2002; 2002US-0349182P.

PR 17-JAN-2002; 2002US-0349733P.

PR 18-JAN-2002; 2002US-0350263P.

PR 24-JAN-2002; 2002US-0351977P.

PR 28-MAY-2002; 2002US-0383758P.

PR 05-JUN-2002; 2002US-0385969P.

PR 11-JUN-2002; 2002US-0387834P.

PR 17-JUL-2002; 2002US-0396407P.

PR 30-SEP-2002; 2002US-0415115P.

PR 03-JAN-2003; 2003US-00336603.

XX (CURA-) CURAGEN CORP.

XX Grose WM, Alsobrook JP, Anderson DW, Burgess CE, Edinger SR;

PI Ellerman K, Furtak K, Gangolli EA, Gerlach VL, Gilbert JA;

PI Gunther E, Gorman L, Guo X, Ji W, Li L, Miller CE, Padigar M;

PI Patturajan M, Rastelli L, Macdougall JR, Mishra VS, Smithson G;

PI Spytek KA, Stone DJ, Shenoy SG, Taupier RJ, Vernet CAM, Zhong M;

PI Malyankar UM, Millet I, Kekuda R;

XX WPI; 2003-587288/55.

DR N-PSDB; ADD49104.

XX New isolated NOVX polypeptides and polynucleotides, useful for
 PT preventing, diagnosing or treating NOVX-associated disorders, e.g.
 PT osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
 PT asthma, or infections.

PS Claim 1; Page 181-182; 31pp; English.

XX The present invention relates to novel NOV proteins and their coding
 CC sequences (ADD49028-ADD49131). The proteins and coding sequences are
 CC useful in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, preferably a NOV-associated disorder
 CC such as metabolic disorders, diabetes, obesity, infectious diseases
 CC (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer,

CC cardiovascular diseases (hypertension, atherosclerosis),
 CC neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
 CC epilepsy, immune disorders (osteoarthritis), hematopoietic disorders,
 CC inflammatory skin disorders, asthma and various dyslipidemias. The coding
 CC sequences and proteins may also be used as targets for the identification
 CC of small molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, in gene therapy, in generation of antibodies that bind
 CC immunospecifically to NOV substances for use in therapeutic or diagnostic
 CC methods.
 XX
 XX Sequence 530 AA;

Query Match 63.1%; Score 41; DB 7; Length 530;
 Best Local Similarity 75.0%; Pred. No. 2.9e+02;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LPAVVGSLSPGEQ 12
 |||||
 Db 416 LPFVVGSPAEQ 427

RESULT 14

ADD49091
 ID ADD49091 standard; protein; 549 AA.

XX
 AC ADD49091;

XX
 DT 15-JAN-2004 (first entry)

XX
 DE Human NOV15d SEQ ID 64.

XX Antidiabetic; anorectic; cardiatic; hypotensive; antiarteriosclerotic;
 KW virucide; antibacterial; fungicide; protozoacide; nootropic;
 KW neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;
 KW antiarthritic; antiinflammatory; dermatological; antiasthmatic;
 KW antilipemic; gene therapy; NOV protein; metabolic disorder; diabetes;
 KW obesity; viral infection; bacterial infection; fungal infection;
 KW helminthic infection; protozoal infection; anorexia; cancer;
 KW cardiovascular disease; hypertension; atherosclerosis;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;
 KW inflammatory skin disorder; asthma; dyslipidemia; human.

XX Homo sapiens.

XX
 PN WO2003060149-A2.

XX
 PD 24-JUL-2003.

XX
 XF 06-JAN-2003; 2003WO-US000252.

XX
 PR 04-JAN-2002; 2002US-0345222P.

PR 14-JAN-2002; 2002US-0348693P.

PR 16-JAN-2002; 2002US-0349182P.

PR 17-JAN-2002; 2002US-0349733P.

PR 18-JAN-2002; 2002US-0350263P.

PR 24-JAN-2002; 2002US-0351977P.

PR 28-MAY-2002; 2002US-0383758P.

PR 05-JUN-2002; 2002US-0385969P.

PR 11-JUN-2002; 2002US-0387834P.

PR 17-JUL-2002; 2002US-0396407P.

PR 30-SEP-2002; 2002US-0415115P.

PR 03-JAN-2003; 2003US-00336603.

XX (CURA-) CURAGEN CORP.

XX
 XX Groese WM, Alsobrook JP, Anderson DW, Burgess CE, Edinger SR;
 PI Ellerman K, Furtak K, Gangolli EA, Gerlach VL, Gilbert JA;
 PI Gunther E, Gorman L, Guo X, Ji W, Li L, Miller CE, Padigaru M;
 PI Patturajan M, Rastelli L, Macdougall JR, Mishra VS, Smithson G;
 PI Spytek KA, Stone DJ, Shenoy SG, Taupier RJ, Vernet CM, Zhong M;
 PI Malyankar UM, Millet I, Kekuda R;

XX WPI; 2003-587288/55.
 DR N-PSDB; ADD49090.
 XX
 XX New isolated NOVX polypeptides and polynucleotides, useful for
 PT preventing, diagnosing or treating NOVX-associated disorders, e.g.
 PT osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,
 PT asthma, or infections.
 XX
 PS Claim 1; Page 177; 311pp; English.

XX The present invention relates to novel NOV proteins and their coding
 CC sequences (ADD49028-ADD49131). The proteins and coding sequences are
 CC useful in the manufacture of a medicament for treating a syndrome
 CC associated with a human disease, preferably a NOV-associated disorder
 CC such as metabolic disorders, diabetes, obesity, infectious diseases
 CC (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer,
 CC cardiovascular diseases (hypertension, atherosclerosis),
 CC neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
 CC epilepsy, immune disorders (osteoarthritis), hematopoietic disorders,
 CC inflammatory skin disorders, asthma and various dyslipidemias. The coding
 CC sequences and proteins may also be used as targets for the identification
 CC of small molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, in gene therapy, in generation of antibodies that bind
 CC immunospecifically to NOV substances for use in therapeutic or diagnostic
 CC methods.
 XX
 XX Sequence 549 AA;

Query Match 63.1%; Score 41; DB 7; Length 549;
 Best Local Similarity 75.0%; Pred. No. 3e+02;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LPAVVGSLSPGEQ 12
 |||||
 Db 419 LPFVVGSPAEQ 430

RESULT 15

ABG70277
 ID ABG70277 standard; protein; 579 AA.

XX
 AC ABG70277;

XX
 DT 05-NOV-2002 (first entry)

XX
 DE Human Glypican-2 Precursor-like protein #1.

XX Human; NOVX; pathological condition; NOVX-associated disorder;
 KW Von Hippel-Lindau syndrome; cirrhosis; transplantation disorder;
 KW pancreatitis; obesity; diabetes; autoimmune disease; infertility;
 KW renal artery stenosis; interstitial nephritis; glomerulonephritis;
 KW polycystic kidney disease; cataract; Alzheimer's disease; cancer;
 KW acoustic trauma; cardiomyopathy; atherosclerosis; hypertension;
 KW congenital heart defect; scleroderma; endometriosis; haemophilia;
 KW dementia; stroke; Parkinson's disease; Huntington's disease; epilepsy;
 KW multiple sclerosis; anxiety; pain; leukaemia; hypothyroidism; psoriasis;
 KW acne; wound; asthma; human disease; calpain; epsin; zinc finger;
 KW low density lipoprotein B; LDLB; purinoceptor; CG8841; synaptotagmin;
 KW serine protease TLRP; mitogen activated protein kinase kinase-2;
 KW glypican-2 precursor; thymosin beta-10.

XX Homo sapiens.

XX
 PN WO200255702-A2.

XX
 PD 18-JUL-2002.

XX
 PF 26-OCT-2001; 2001WO-US050925.

XX
 PR 26-OCT-2000; 2000US-0243320P.

XX
 PR 26-OCT-2000; 2000US-0243592P.

PR 26-OCT-2000; 2000US-0243642P.
 PR 27-OCT-2000; 2000US-0243681P.
 PR 27-OCT-2000; 2000US-0243663P.
 PR 31-OCT-2000; 2000US-0244443P.
 PR 01-NOV-2000; 2000US-0244995P.
 PR 01-NOV-2000; 2000US-0245029P.
 PR 02-NOV-2000; 2000US-0245293P.
 PR 02-NOV-2000; 2000US-0245315P.
 PR 02-NOV-2000; 2000US-0245316P.
 PR 19-JAN-2001; 2001US-0262994P.
 PR 15-FEB-2001; 2001US-0269056P.
 PR 02-MAR-2001; 2001US-0272923P.
 PR 15-MAR-2001; 2001US-0276565P.
 PR 07-SEP-2001; 2001US-0318119P.
 XX (CURA-) CURAGEN CORP.
 PA
 XX Gangolli EA, Spytek KA, Gilbert J, Casman S, Blalock A, Li L;
 PI Vernet CAM, Shency S, Mishra V, Furtak K, Gerlach V, Edinger S;
 PI Malyankar U, Stone D, Millet I, Smithson G, Gunther E, Padigaru M;
 PI Taupier RJ, Anderson D;
 XX
 DR WPI; 2002-590673/63.
 DR N-PSDB; ABK51684.
 XX
 PT Isolated NOVX polypeptides and nucleic acid molecules useful for
 PT treating, preventing, diagnosing and researching pathological conditions
 PT in humans with a NOVX-associated disorders, e.g. cancer, stroke or
 PT Alzheimer's disease.
 XX
 PS Claim 1; Page 57; 236pp; English.
 XX
 CC The present invention relates to a new polypeptide that comprises any of
 CC 17 fully defined sequences of 43-990 amino acids given in the
 CC specification. The NOVX polypeptide, nucleic acid and antibody of the
 CC invention are useful for treating or preventing a pathological condition
 CC in humans with a NOVX-associated disorder, e.g. Von Hippel-Lindau
 CC syndrome, cirrhosis, transplantation disorders, pancreatitis, obesity,
 CC diabetes, autoimmune disease, renal artery stenosis, interstitial
 CC nephritis, glomerulonephritis, polycystic kidney disease, cataract,
 CC Alzheimer's disease, acoustic trauma, cancer, infertility, heart
 CC cardiomyopathies, atherosclerosis, hypertension, congenital heart
 CC defects, scleroderma, endometriosis, haemophilia, dementia, stroke,
 CC Parkinson's disease, Huntington's disease, epilepsy, multiple sclerosis,
 CC anxiety, pain, leukaemia, hypothyroidism, psoriasis, acne, wounds and
 CC asthma. They are also useful for the manufacture of a medicament for
 CC treating a syndrome associated with a human disease, specifically a NOVX-
 CC associated disorder. They may also be useful in therapeutic applications
 CC including protein therapy, as small molecule drug targets, as antibody
 CC targets, as diagnostic and/or prognostic markers, in gene therapy, as
 CC research tools and in tissue regeneration. The present amino acid
 CC sequence represents one of the 17 novel proteins of the invention
 XX
 SQ Sequence 579 AA;

Query Match 63.1%; Score 41; DB 5; Length 579;
 Best Local Similarity 75.0%; Pred No. 3.2e+02;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 LPAVVGLSPEEQ 12
 |||||
 Db 439 LPPVVGGSPEAQ 450

Search completed: May 7, 2004, 12:33:46
 Job time : 36.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:20:30 ; Search time 5.4 Seconds
(without alignments)
144.639 Million cell updates/sec

Title: US-09-786-214A-15

Perfect score: 75

Sequence: 1 AGLFAVGLSPGEQE 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	57.3	658	1 VG18_BPT4	P13332 bacteriophage
2	42	56.0	326	1 MER_METTI	Q9uxp0 methanolobu
3	42	56.0	508	1 COBI_MYCTU	Q10677 mycobacteri
4	41	54.7	156	1 RUVX_CAUCR	Q9a5k8 caulobacter
5	41	54.7	213	1 MDCG_XANAC	Q9ppw9 xanthomonas
6	41	54.7	520	1 AT15_YEAST	P25641 saccharomyc
7	41	54.7	774	1 LOL2_HUMAN	Q9v4k0 homo sapien
8	40	53.3	157	1 RISB_PYRPU	Q8u4l8 pyrococcus
9	40	53.3	359	1 ALF2_PEA	P42527 psium sativ
10	40	53.3	359	1 ALF2_PEA	Q65735 ciccer ariet
11	40	53.3	361	1 VAL1_TMOV	Q06657 tomato mott
12	40	53.3	633	1 PLB5_SCHPO	Q9y7n6 schizosacch
13	40	53.3	773	1 YHGF_ECOLI	P46837 escherichia
14	40	53.3	813	1 CADM_MOUSE	Q9wtp5 mus musculu
15	40	53.3	2364	1 PCGA_BOVIN	P13608 bos taurus
16	39	52.0	115	1 KV31_HUMAN	P04433 homo sapien
17	39	52.0	115	1 KV51_MOUSE	P01642 mus musculu
18	39	52.0	146	1 DTD_XANAC	Q9pg32 xanthomonas
19	39	52.0	385	1 YG53_METUA	Q59047 methanococc
20	39	52.0	597	1 NR41_RAT	P22829 rattus norv
21	39	52.0	1402	1 N160_MOUSE	Q9z0w3 mus musculu
22	39	52.0	1596	1 GLI3_HUMAN	P10071 homo sapien
23	39	52.0	1636	1 RUD3_YEAST	P25558 saccharomyc
24	38	50.7	146	1 DTD_XANCP	Q9p4h1 xanthomonas
25	38	50.7	299	1 DAF1_STRCO	O86841 streptomyce
26	38	50.7	299	1 NUCG_BOVIN	P38447 bos taurus
27	38	50.7	331	1 ILVC_ANASP	Q8yum5 anabaena sp
28	38	50.7	344	1 ILVC_OCELI	Q9en66 oceanobacil
29	38	50.7	390	1 COBL_MYCTU	Q10671 mycobacteri
30	38	50.7	395	1 HNEB_ARCFU	O29750 archaeoglob
31	38	50.7	429	1 R31_LEULA	P50889 leuconostoc
32	38	50.7	484	1 COBQ_PSEDE	P29932 pseudomonas
33	38	50.7	507	1 CATA_PICAN	P30263 pichia angu

34	38	50.7	516	1 C4AD_DROME	Q9v4t3 drosophila
35	38	50.7	628	1 HNFA_MOUSE	P22361 mus musculu
36	38	50.7	628	1 HNFA_RAT	P15257 rattus norv
37	38	50.7	631	1 HNFA_HUMAN	P20823 homo sapien
38	38	50.7	699	1 EFG_HAEIN	P43925 haemophilus
39	38	50.7	700	1 EFG_PASMO	P57938 pasteurella
40	38	50.7	876	1 TOF1_VIBCH	Q9krb2 vibrio chol
41	38	50.7	895	1 ANDR_MACFA	O97952 macaca fasc
42	38	50.7	895	1 ANDR_PAPHA	O97960 papio hamad
43	38	50.7	899	1 ANDR_MOUSE	P19091 mus musculu
44	38	50.7	902	1 ANDR_RAT	P15207 rattus norv
45	38	50.7	907	1 ANDR_CANFA	Q9tt90 canis faml

ALIGNMENTS

RESULT 1
VG18_BPT4
ID VG18_BPT4 STANDARD; PRT; 658 AA.
AC P13332; Q9TOU3;
DT 01-JAN-1990 (Rel. 13, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tail sheath protein Gp18.
GN 18.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10665;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=88155753; PubMed=2964531;
RA Arisaka F., Nakako T., Takahashi H., Ishii S.-I.;
RT "Nucleotide sequence of the tail sheath gene of bacteriophage T4 and
RT amino acid sequence of its product."
RL J. Virol. 62:1186-1193 (1988).
RN [2]
SEQUENCE FROM N.A.
RP MEDLINE=22514363; PubMed=12626685;
RA Miller E.S., Kutter E., Mosig G., Arisaka F., Kunisawa T., Ruger W.;
RT "Bacteriophage T4 genome."
RL Microbiol. Mol. Biol. Rev. 67:86-156 (2003).
RN [3]
SEQUENCE OF 638-658 FROM N.A.
RA Arisaka F., Ishimoto L., Kasaavetis G., Kumazaki T., Ishii S.-I.;
RT "Nucleotide sequence of the tail tube structural gene of
RT bacteriophage T4."
RL J. Virol. 62:882-886 (1988).
CC -!- FUNCTION: The contractile tail of bacteriophage T4 consists of a
CC contractile sheath, a tube and a baseplate. 144 protomers of Gp18,
CC arranged in 24 annuli, form the contractile tail sheath.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: 30% IDENTITY TO THE TAIL SHEATH PROTEIN OF PHAGE P517,
CC AND BY EXTENSION, TO P.AERUGINOSA R-TYPE PYOCINS (ANTIBIOTICS).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M19085; AAA32541.1; -;
DR EMBL; AF158101; AAD42423.1; -;
DR PIR; JF0021; GKBP74.
DR InterPro; IPR007067; Phage sheath 1.
DR Pfam; PF04984; Phage_sheath_1; 1.
DR Structural protein.
FT INIT MET 0 175 SER-RICH.
FT DOMAIN 171 175

```

FT CONFLICT 99 99 D -> E (IN REF. 1).
FT CONFLICT 147 150 GKNY -> AKII (IN REF. 1).
FT CONFLICT 300 300 E -> G (IN REF. 1).
FT CONFLICT 398 398 A -> V (IN REF. 1).
FT CONFLICT 453 453 H -> Y (IN REF. 1).
FT CONFLICT 594 594 N -> I (IN REF. 1).
SQ SEQUENCE 658 AA; 71199 MW; 6E045F40D39AF21D CRC64;

Query Match
Best Local Similarity 57.3%; Score 43; DB 1; Length 658;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPAVVGLSPGE 13
DB 217 GIFGVVALYFGE 228

RESULT 2
MER_METTII
ID MER_METTII STANDARD; PRT; 326 AA.
AC Q9UXP0;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Coenzyme F420-dependent N(5),N(10)-methylentetrahydromethanopterin
DE reductase (EC 1.5.99.11) (Methylene-H(4)MPT reductase).
GN MER OR FFDA.
OS Methanobolus tindarius.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanobolus.
OX NCBI_TaxID=2221;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2278;
RX MEDLINE=99132696; PubMed=9933933;
RA Westenberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA Gottschalk G., Blaut M.;
RA "The F420H2-dehydrogenase from Methanobolus tindarius: cloning of the
RA ffd operon and expression of the genes in Escherichia coli.";
RA FEMS Microbiol. Lett. 170:389-398(1999).
CC -!- FUNCTION: Catalyzes the reversible reduction of methylene-H(4)MPT
CC to methyl-H(4)MPT (By similarity).
CC -!- CATALYTIC ACTIVITY: N(5),N(10)-methylentetrahydromethanopterin +
CC reduced coenzyme F420 = 5-methyl-5,6,7,8-tetrahydromethanopterin +
CC coenzyme F420.
CC -!- PATHWAY: Methanogenesis from carbon dioxide; fifth step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the mer family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AJ011519; CAB56639.1; -.
DR PIR; T45226; T45226.
DR HAMAP; MF 01091; -.
DR InterPro; IPR002103; Bac_luciferase.
DR Pfam; PF00296; bac_luciferase; 1.
KW Methanogenesis; One-carbon metabolism; Oxidoreductase.
SQ SEQUENCE 326 AA; 34043 NW; 16F3AB9733A45D82 CRC64;

Query Match
Best Local Similarity 56.0%; Score 42; DB 1; Length 326;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLPAVVGLSPGE 14
DB 84 SGGRAILGLGFGEQ 97

```

```

RESULT 3
COBI_MYCTTU
ID COBI_MYCTTU STANDARD; PRT; 508 AA.
AC Q10677;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
DE methyltransferase (EC 2.1.1.130) (S-adenosyl-L-methionine-precorrin-2
DE methyltransferase) (SP2MT); Precorrin-3 methylase (EC 2.1.1.-)].
GN COBIJ OR COBI OR RV2066 OR MT2126 OR MTCY49.05 OR MB2092.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=11773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence.";
RA Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RA J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RA Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC PRECORRIN-3A (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
CC adenosyl-L-homocysteine + precorrin-3A.
CC -!- PATHWAY: Cobalamin biosynthesis.
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC AND CBIL/COBI.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; Z73966; CAA98214.1; -.
DR EMBL; AE007063; AAK46406.1; -.
DR EMBL; BX248341; CAD96945.1; -.

```

```

DR PIR: E70764; E70764.
DR TIGR: MT2126; -.
DR TubercuList; RV2066; -.
DR InterPro: IPR006364; Cobi CblL.
DR InterPro: IPR006363; CobiJ.
DR InterPro: IPR000878; Cox/pox Metransf.
DR InterPro: IPR003043; Uropor Metransf.
DR Pfam: PF00590; TP_methylase; 2.
DR TIGRFAMS: TIGR01467; cobi cblI; 1.
DR TIGRFAMS: TIGR01466; cobiJ cblH; 1.
DR PROSITE: PS00839; SUMT 1; 1.
DR PROSITE: PS00840; SUMT 2; 1.
KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
KW Methyltransferase; Multifunctional enzyme; Complete proteome.
FT DOMAIN 1 243 PRECORRIN-2 C20-METHYLTRANSFERASE.
FT DOMAIN 244 508 PRECORRIN-3 METHYLASE.
SQ SEQUENCE 508 AA; 53910 MW; 95AC066F022C4DC1 CRC64;

Query Match
Best Local Similarity 56.0%; Score 42; DB 1; Length 508;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 GLPAVVGLSPGQEE 15
Db 247 GTVAVVGLGFGSD 260

RESULT 4
RUVX CAUCR
ID RUVX CAUCR STANDARD; PRT; 156 AA.
AC Q9A5K8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative Holliday junction resolvase (EC 3.1.1.1).
GN CC2439.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 39089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.B., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Ermolaeva M., White O.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Fraser C.M.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001)
CC -!- FUNCTION: Could be a nuclease that resolves Holliday junction
CC intermediates in genetic recombination.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the ygfH HJR family.
CC
CC
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; AE005913; AA024410.1; -.
CC PIR; F87551; F87551.
CC TIGR; CC2439; -.
CC HAMAP; MF 00651; -.
CC InterPro: IPR005227; Cons hypoth250.
CC InterPro: IPR006641; YgfF.
CC Pfam; PF03652; UPP0081; 1.

```

```

DR SMART: SM00732; YgfF; 1.
DR TIGRFAMS: TIGR00250; TIGR00250; 1.
KW Hydrolyase; Nuclease; DNA repair; DNA recombination; Complete proteome.
SQ SEQUENCE 156 AA; 17142 MW; 21F54D8648396141 CRC64;

Query Match
Best Local Similarity 54.7%; Score 41; DB 1; Length 156;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 AVVGLSPGQEE 14
Db 18 AVVGLDPGEK 27

RESULT 5
MDCG_XANAC
ID MDCG_XANAC STANDARD; PRT; 213 AA.
AC Q8PFW9;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phosphoribosyl-dephospho-CoA transferase (EC 2.7.7.-) (Holo-ACP
DE synthase).
GN MDCG OR XAC0564.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Queglio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A.F., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RA "Comparison of the genomes of two Xanthomonas pathogens with differing
RA host specificities.";
RL Nature 417:459-463(2002)
CC -!- FUNCTION: Transfers 2-(5'-triphosphoribosyl)-3'-
CC dephosphocoenzyme-A to the apo-acyl carrier protein of the
CC malonate decarboxylase to yield holo-acyl carrier protein (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: 2'-(5'-triphosphoribosyl)-3'-dephospho-CoA +
CC apo-ACP = holo-ACP + diphosphate.
CC -!- SIMILARITY: Belongs to the mdcG family.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; AE011683; AA035453.1; -.
CC HAMAP; MF 00650; -.
CC Transferase; Nucleotidyltransferase; Complete proteome.
FT ACT_SITE 135 135 BY SIMILARITY.
FT ACT_SITE 137 137 BY SIMILARITY.
SQ SEQUENCE 213 AA; 23013 MW; C378D8A975C75A17 CRC64;

```

```

Query Match          54.7%; Score 41; DB 1; Length 213;
Best Local Similarity 64.3%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGLPAVUGLSPGE 14
DB 33 AGLPAVARGDSQ 46

RESULT 6
AT15 YEAST
ID AT15 YEAST STANDARD; PRT; 520 AA.
AC P25641; O8N1L6.
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE Autophagy related putative lipase ATG15 (EC 3.1.1.3) (Cytoplasm to
DE vacuole targeting protein 17)
GN ATG15 OR AUT5 OR CVT17 OR YCR068W OR YCR68W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, GLYCOSYLATION, AND MUTAGENESIS OF
RP SER-332.
RX MEDLINE=21125771; PubMed=11085977;
RA Teter S.A., Eggerton K.P., Scott S.V., Kim J., Fischer A.M.,
RA Klionsky D.J.,
RT "Degradation of lipid vesicles in the yeast vacuole requires function
RT of Cvt17, a putative lipase.";
RL J. Biol. Chem. 276:2083-2087(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=92243356; PubMed=1574125;
RA Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
RA Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
RA Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
RA Biotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
RA Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
RA De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
RA Dujon B., Dueterhoef A., Erdmann D., Esteban M., Fabre F.,
RA Fairhead C.A., Faye G., Feldmann H., Fiers W.,
RA Francouze-Gaillard M.-C., Franco L., Frontali L., Fukuhara H.,
RA Fuller L.J., Gent M.B., Gigot D., Gilliquet V., Glansdorff N.,
RA Goffeau A., Grenson M., Grisanti P., Griwell L.A., Haasemann M.,
RA Hatat D., Hegemann J.H., Herbert C.J., Hilger F., Hohmann S.,
RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Isono K., Jackman P.,
RA Jacq C., Jacquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,
RA Kleinhaus U., Kreis P., Lairanchi G., Lewis C., van der Linden C.G.,
RA Lucchini G., Lutzenkirchen K., Maat C., Mannhaupt G., Manzano M.E.,
RA Messenguy F., Mewes H.-W., Molemans F., McConnell D., McKee R.A.,
RA Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
RA Perea J., Philippsen P., Pierard A., Planta R.J., Plevani P.,
RA Poetsch B., Söhl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
RA Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
RA Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
RA Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
RA Spiegelberg R., Staveva L.I., Steensma H.Y., Steiner S., Thierry A.,
RA Threes G., Triano L.N., Urrestazu L.A., Valle G., Vetter I.,
RA von Vliet-Reedijk J.C., Voickaert G., Vreken P., Warrington J.R.,
RA Zimmermann F.K., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
RT "The complete DNA sequence of yeast chromosome III.";
RL Nature 357:38-46(1992).
RN [3]
RP REVISIONS TO C-TERMINUS.
RA Valles G., Voickaerts G.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP FUNCTION, SUBCELLULAR LOCATION, GLYCOSYLATION, AND MUTAGENESIS OF

```

```

RP SER-332.
RX MEDLINE=21450820; PubMed=11566994;
RA Eppie U.D., Suriapranata I., Eskelinen E.-L., Thumm M.;
RT "Aut5/Cvt17p, a putative lipase essential for disintegration of
RT autophagic bodies inside the vacuole.";
RL J. Bacteriol. 183:5942-5955(2001).
RN [5]
RP FUNCTION, SUBCELLULAR LOCATION, TOPOLOGY, AND GLYCOSYLATION.
RX MEDLINE=22499636; PubMed=12499386;
RA Eppie U.D., Eskelinen E.-L., Thumm M.;
RT "Intravacuolar membrane lysis in Saccharomyces cerevisiae. Does
RT vacuolar targeting of Cvt17/Aut5p affect its function?";
RL J. Biol. Chem. 278:7810-7821(2003).
CC -!- FUNCTION: Essential for lysis of subvacuolar cytoplasm to vacuole
CC targeted bodies and intravacuolar autophagic bodies. Involved in
CC the lysis of intravacuolar multivesicular body (MVB) vesicles.
CC -!- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
CC fatty acid anion.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum. From ER, targeted to vacuolar lumen at the MVB vesicles
CC via the Golgi and the prevacuolar compartment (PVC).
CC -!- PTM: Glycosylated.
CC -!- SIMILARITY: Belongs to the AB hydrolase superfamily. Lipase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59720; CAC42987.1; -.
CC PIR; S19483; S19483.
CC PIR; S74292; S74292.
CC GerOnline; 138970; -.
CC SGD; S0000664; ATG15.
CC GO; GO:0016021; C: integral to membrane; IPI.
CC GO; GO:0005775; C: vacuolar lumen; IDA.
CC GO; GO:0016298; F: lipase activity; IMP.
CC GO; GO:0006914; P: autophagy; IDA.
CC GO; GO:0030397; P: membrane degradation; IDA.
CC GO; GO:0006624; P: vacuolar protein processing/maturation; IDA.
CC InterPro; IPR002921; Lipase_3.
CC InterPro; IPR008262; Lipase_AS.
CC Pfam; PF01764; Lipase_3; 1.
CC PROSITE; PS00120; LIPASE_SER; 1.
CC Hydrolase; Lipid degradation; Autophagy; Endoplasmic reticulum;
CC Signal-anchor; Transmembrane; Glycoprotein.
CC DOMAIN 1 14 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN).
CC TRANSMEM 15 35 LUMENAL.
CC DOMAIN 36 520
CC ACT SITE 332 332 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 202 202 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 208 208 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT MUTAGEN 332 332 S-A: LOSS OF FUNCTION.
CC SQ SEQUENCE 520 AA; 58435 MW; B56ABAB72B999019 CRC64;

Query Match          54.7%; Score 41; DB 1; Length 520;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPAVUGLSPGE 13
DB 346 GLPAVAFESPE 357

RESULT 7
LOL2 HUMAN
ID LOL2 HUMAN STANDARD; PRT; 774 AA.
AC Q9Y4K0; Q9BW70; Q9Y5Y8;

```

16-OCT-2001 (Rel. 40, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lysyl oxidase homolog 2 precursor (EC 1.4.3.-) (Lysyl oxidase-like
 protein 2) (Lysyl oxidase related protein 2) (Lysyl oxidase-related
 protein WS9-14).
 GN LOXL2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 EX MEDLINE=97236759; PubMed=9079631;
 RT Saito H., Papaconstantinou J., Sato H., Goldstein S.;
 RT "Regulation of a novel gene encoding a lysyl oxidase-related protein
 in cellular adhesion and senescence.";
 RL J. Biol. Chem. 272:8157-8160(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grinstead J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smaluk D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Makra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 202-774 FROM N.A.
 RC TISSUE=Placenta, and Spleen;
 RX MEDLINE=99230328; PubMed=10212285;
 RA Jourdan-Le Saux C., Tronecker H., Bogic L., Bryant-Greenwood G.D.,
 RA Boyd C.D., Geislar K.;
 RT "The LOXL2 gene encodes a new lysyl oxidase-like protein and is
 expressed at high levels in reproductive tissues.";
 RL J. Biol. Chem. 274:12939-12944(1999).
 CC -1- COFACTOR: Copper and LTQ (By similarity).
 CC -1- SUBCELLULAR LOCATION: Extracellular (Potential).
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues. Highest expression
 CC in reproductive tissues, placenta, uterus and prostate.
 CC -1- PTM: The lysine tyrosylquinone cross-link (LTQ) is generated by
 CC condensation of the epsilon-amino group of a lysine with a
 CC topaquinone produced by oxidation of tyrosine.
 CC -1- SIMILARITY: Contains 4 SRCR domains.
 CC -1- SIMILARITY: Belongs to the lysyl oxidase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U89942; AAB49697.1; -;
 CC EMBL; BC000594; AAB49697.1; -;
 CC EMBL; AF117949; AAB49697.1; -;
 CC Genew; HGNC:6666; LOXL2.

DR MIM; 606663; -;
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR GO; GO:0005489; F:electron transporter activity; TAS.
 DR GO; GO:0007568; P:aging; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR GO; GO:0006464; P:protein modification; TAS.
 DR InterPro; IPR001895; Lysyl oxidase.
 DR Pfam; PF01186; Lysyl oxidase; 1.
 DR Pfam; PF00530; SRCR; 4.
 DR PRINTS; PR00074; LYSYLOXIDASE.
 DR ProDom; PD013887; Lysyl oxidase; 1.
 DR SMART; SM00202; SR; 4.
 DR PROSITE; PS00926; LYSYL OXIDASE; 1.
 DR PROSITE; PS00420; SRCR_1; 1.
 DR PROSITE; PS0287; SRCR_2; 4.
 KW Oxidoreductase; Copper; Glycoprotein; Repeat; Signal; LTO.
 FT SIGNAL 1 25
 FT CHAIN 26 774
 FT DOMAIN 58 159
 FT DOMAIN 188 302
 FT DOMAIN 326 425
 FT DOMAIN 435 551
 FT DOMAIN 548 751
 FT METAL 626 626
 FT METAL 628 628
 FT METAL 630 630
 FT CROSSLINK 653 689
 FT MOD_RES 689 689
 FT CARBOHYD 288 288
 FT CARBOHYD 455 455
 FT CARBOHYD 644 644
 FT CONFLICT 295 295
 FT CONFLICT 570 570
 FT CONFLICT 652 652
 FT CONFLICT 746 746
 FT CONFLICT 774 774
 SQ SEQUENCE 774 AA; 86724 MW; 9DF5D25D4824BCDC CRC64;
 Query Match 54.7%; Score 41; DB 1; Length 774;
 Best Local Similarity 66.7%; Pred. No. 83;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 GLPAVGLSPGE 13
 Db 294 GLPAVWCVPEQ 305
 RESULT 8
 RISB PYRFU STANDARD; PRT; 157 AA.
 AC Q8U4L8; 2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE 6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.9) (DMRL synthase)
 DE (lumazine synthase) (Riboflavin synthase beta chain).
 GN RIBH OR PF0063.
 OS Pyrococcus furiosus.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=2261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the Pyrococcus furiosus genome.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex

CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
 CC ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-
 CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
 CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-
 CC amino-2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-
 CC butanone-4-phosphate yielding 6,7-dimethyl-8-lumazine (By
 CC similarity).
 CC -!- CATALYTIC ACTIVITY: 2,6,7-dimethyl-8-(1-D-ribityl)lumazine =
 CC riboflavin + 4-(1-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
 CC -!- PATHWAY: Riboflavin biosynthesis; last step.
 CC -!- SIMILARITY: Belongs to the DMRL synthase family.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB010132; AAL80187.1; -.
 CC HAMAP; MF 00178; -; 1.
 CC InterPro; IPR002180; DMRL synthase.
 CC Pfam; PF00885; DMRL synthase; 1.
 CC ProDom; PD003664; DMRL synthase; 1.
 CC TIGRFAMs; TIGR00114; rfbB; 1.
 CC Riboflavin biosynthesis; Transferase; Complete proteome.
 CC SEQUENCE 157 AA; 17517 MW; 34BB6E720FD20C1 CRC64;
 SQ
 CC
 CC Query Match 53.3%; Score 40; DB 1; Length 157;
 CC Best Local Similarity 42.9%; Pred. No. 25;
 CC Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 CC
 CC QY 2 GLPVAVGLSPGQE 15
 CC Db 107 GIPVIGIVPADE 120
 CC
 CC RESULT 9
 CC ALF2_PEA
 CC ID ALF2_PEA STANDARD; PRT; 359 AA.
 CC AC P46257;
 CC DT 01-NOV-1995 (Rel. 32, Created)
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Fructose-bisphosphate aldolase, cytoplasmic isozyme 2 (EC 4.1.2.13).
 CC Pisum sativum (Garden pea).
 CC OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 CC OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
 CC OX NCBI_TaxID=3888;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Leaf;
 CC RA Pelzer-Reith B., Schnarrenberger C.;
 CC RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycero-
 CC phosphate + D-glyceraldehyde 3-phosphate.
 CC -!- PATHWAY: Glycolysis; sixth step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to Class I fructose-bisphosphate aldolase
 CC family.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X89829; CAA61947.1; -.
 CC PIR; S58167; S58167.

DR HSP; P00883; IADO.
 DR InterPro; IPR000741; Aldolase I.
 DR Pfam; PF00274; Glycolytic_enz; 1.
 DR ProDom; PD001128; Aldolase I; 1.
 DR PROSITE; PS00158; ALDOLASE CLASS I; 1.
 KW Lyase; Schiff base; Glycolysis; Multigene family.
 FT BINDING 52
 FT C-1-PHOSPHATE GROUP OF THE SUBSTRATE
 FT (BY SIMILARITY).
 FT BINDING 143 143
 FT C-1-PHOSPHATE GROUP OF THE SUBSTRATE
 FT (BY SIMILARITY).
 FT BINDING 226 226
 FT SCHIFF-BASE WITH DIHYDROXYACETONE-P
 FT (BY SIMILARITY).
 FT ACT_SITE 359 359
 FT ESSENTIAL FOR ENHANCED ACTIVITY OF THE
 FT ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
 FT AS COMPARED WITH FRUCTOSE 1-PHOSPHATE
 FT (BY SIMILARITY).
 SQ SEQUENCE 359 AA; 38490 MW; C0CAB16E9CC1B9EF CRC64;
 CC
 CC Query Match 53.3%; Score 40; DB 1; Length 359;
 CC Best Local Similarity 60.0%; Pred. No. 56;
 CC Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC
 CC QY 1 AGLPAVGLSPGQE 15
 CC Db 258 AAVPAVFLSGQSE 272
 CC
 CC RESULT 10
 CC ALF_CICAR
 CC ID ALF_CICAR STANDARD; PRT; 359 AA.
 CC AC O65735;
 CC DT 15-DEC-1998 (Rel. 37, Created)
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Fructose-bisphosphate aldolase, cytoplasmic isozyme (EC 4.1.2.13).
 CC GN ALDC.
 CC OS Cicer arietinum (Chickpea) (Garbanzo).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 CC OC eurosids I; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.
 CC OX NCBI_TaxID=3827;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=cv. Castellana; TISSUE=Etioolated epicotyl;
 CC RA Dopico B., Munoz F.J., Labrador E.;
 CC RT "cDNA and deduced amino-acid sequence of a cytosolic aldolase from
 CC Cicer arietinum L. epicotyls";
 CC RL (in) Plant Gene Register PGR98-110.
 CC CC -!- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycero-
 CC phosphate + D-glyceraldehyde 3-phosphate.
 CC -!- PATHWAY: Glycolysis; sixth step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to Class I fructose-bisphosphate aldolase
 CC family.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ005041; CAA06308.1; -.
 CC HSP; P00883; IADO.
 CC InterPro; IPR000741; Aldolase I.
 CC Pfam; PF00274; Glycolytic_enz; 1.
 CC ProDom; PD001128; Aldolase I; 1.
 CC PROSITE; PS00158; ALDOLASE CLASS I; 1.
 KW Lyase; Schiff base; Glycolysis.
 FT BINDING 52 52
 FT C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
 FT BINDING 142 142
 FT C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
 FT BINDING 226 226
 FT SCHIFF-BASE WITH DIHYDROXYACETONE-P.

```

FT ACT_SITE 359 359 ESSENTIAL FOR ENHANCED ACTIVITY OF THE
FT ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
FT AS COMPARED WITH FRUCTOSE 1-PHOSPHATE.
SQ SEQUENCE 359 AA; 38451 MW; DD68864B745A5195 CRC64;

Query Match
Best Local Similarity 60.0%; Pred. No. 56; Length 359;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLPAVGLSPGEQ 15
Db 258 AAVPAVGLSPGEQ 272

RESULT 11
VAL1_TM0V STANDARD; PRT; 361 AA.
ID VAL1_TM0V
AC Q06567;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE AL1 protein.
GN AL1.
OS Tomato mottle virus (isolate Florida) (TM0V).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=36449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93107858; PubMed=1469361;
RA Abouzid A.M., Polston J.E., Hiebert E.;
RT "The nucleotide sequence of tomato mottle virus, a new geminivirus
RT isolated from tomatoes in Florida.";
RL J. Gen. Virol. 73:3225-3229 (1992).
CC -1- SIMILARITY: Belongs to the geminiviruses AL1 protein family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L14460; AAC32414.1; --
CC PIR; JQ1870; JQ1870.
CC InterPro; IPR001191; Gemini_AL1.
CC Pfam; PF00799; Gemini_AL1; 1.
CC PRINTS; PR00227; GEMCOAT1.
CC ProDom; PD000736; Gemini_AL1; 1.
CC ATP-binding.
FT NP_BIND 222 229 ATP (BY SIMILARITY).
SQ SEQUENCE 361 AA; 40516 MW; 813B865CEBAC6950 CRC64;

Query Match
Best Local Similarity 58.3%; Score 40; DB 1; Length 361;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLPAVGLSPGE 13
Db 299 GIPAVGLCPNFE 310

RESULT 12
PLB5_SCHPO STANDARD; PRT; 633 AA.
AC Q9Y7N6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative lysophospholipase C1450.09c precursor (EC 3.1.1.5)
DE (Phospholipase B).
GN SPCC1450.09c.
OS Schizosaccharomyces pombe (Fission yeast).

```

```

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21849401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Kieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moeati D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Sipakowski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880 (2002).
CC -1- FUNCTION: Catalyzes the release of fatty acids from
CC lysophospholipids (By similarity).
CC -1- CATALYTIC ACTIVITY: 2-lysophosphatidylcholine + H(2)O =
CC glycerophosphocholine + a fatty acid anion.
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- SIMILARITY: Belongs to the lysophospholipase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL049559; CAB40176.2; --
CC GeneDB SPombe; SPCC1450.09c; --
CC InterPro; IPR002642; PLAC.
CC Pfam; PF01735; PLA2_B; 1.
CC SMART; SM00022; PLAC; 1.
CC Hypothetical protein; Lipid degradation; Hydrolase; Glycoprotein;
CC Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 633 PUTATIVE LYSOPHOSPHOLIPASE C1450.09c.
FT CARBOHYD 118 118 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 508 508 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).

```



```

FT CARBOHYD 537 537 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 633 AA; 68292 MW; 4987182955893D19 CRC64;

Query Match 53.3%; Score 40; DB 1; Length 633;
Best Local Similarity 75.0%; Pred. No. 97;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 PAVVGLSPGQE 15
Db 76 PASDGLSTGQE 87

RESULT 13
YHGF_ECOLI STANDARD; PRT; 773 AA.
AC P46837; P76689;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein yHGF.
GN YHGF OR B3407.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaser J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=99420866; PubMed=10493123;
RA Fountoulakis M., Takacs M.-P., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195(1999).
CC -!- SIMILARITY: STRONG, TO H.INFLUENZAE HI0568.
CC -!- SIMILARITY: Contains 1 SI motif domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U18997; AAA58204.1; ALT_FRAME.
DR EMBL; U18997; AAA58205.1; ALT_FRAME.
DR EMBL; AE000416; AAC76432.1; ALT_INIT.
DR HSSP; P05055; ISRO.
DR Ecogen; EGI2932; yHGF.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR InterPro; IPR003029; S1.
DR InterPro; IPR006641; YqgFc.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00316; S1; 1.
DR SMART; SM00732; YqgFc; 1.
DR PROSITE; PS0126; S1; 1.
KW RNA-binding; Complete proteome.
FT DOMAIN 651 720 S1 MOTIF.
FT CONFLICT 754 755 QP -> HA (IN REF. 1; AAA58205).
SQ SEQUENCE 773 AA; 85119 MW; EA54D9ED952A8229 CRC64;

```

```

Query Match 53.3%; Score 40; DB 1; Length 773;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGUPAVVGLSPG 12
Db 324 AGURATMGIDPG 335

RESULT 14
CADM_MOUSE STANDARD; PRT; 813 AA.
AC Q9WTE5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cadherin-22 precursor (PB-cadherin).
GN CDH22.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=99326347; PubMed=10398531;
RA Kitajima K., Koshimizu U., Nakamura T.;
RT "Expression of a novel type of classic cadherin, PB-cadherin in
RT developing brain and limb buds.";
RL Dev. Dyn. 215:206-214(1999).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. PB-cadherins may have a role
CC in the morphological organization of pituitary gland and brain
CC tissues.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Predominantly expressed in brain. Abundant in
CC olfactory bulb, cerebrum, and cerebellum, less in pons, medulla,
CC and spinal cord. Low expression in heart. No expression in lung,
CC liver, spleen, kidney, testis, stomach, intestine, colon, and
CC placenta.
CC -!- DEVELOPMENTAL STAGE: Expressed at 9.5 dpc onwards. At 10.5 dpc, in
CC brain (telencephalic vesicles and isthmus), spinal cord and limb
CC buds (in the zone of polarizing activity). At 14.5 dpc, in
CC olfactory bulb and cerebellum.
CC -!- INDUCTION: Down-regulated by thyroid hormone.
CC -!- SIMILARITY: Contains 5 cadherin domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB019618; BAA34426.1; -.
DR MGD; MGI:1341843; Cdh22.
DR HSSP; P15116; LNCJ.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 2.
DR PROSITE; PS0268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 813 CADHERIN-22.
FT DOMAIN 33 621 EXTRACELLULAR (POTENTIAL).

```

FT TRANSMEM 622 642 POTENTIAL.
 FT DOMAIN 643 813 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 61 165 CADHERIN 1.
 FT DOMAIN 166 274 CADHERIN 2.
 FT DOMAIN 275 391 CADHERIN 3.
 FT DOMAIN 392 495 CADHERIN 4.
 FT DOMAIN 496 613 CADHERIN 5.
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 463 463 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 609 609 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 813 AA; 86021 MW; 5510F9848D976567 CRC64;
 Query Match 53.3%; Score 40; DB 1; Length 813;
 Best Local Similarity 60.0%; Pred. No. 1.2e+02;
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 AGLPAVVGLSPEOE 15
 DB 34 ASTPAPSSLSPGAQE 48
 RESULT 15
 ID PGCA BOVIN STANDARD; PRT: 2364 AA.
 AC P13608; P79117; Q28159;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSCP).
 DE protein) (CSCP).
 GN AGC1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 ON NCBI_TaxID=9913;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Hering T.M., Kollar J., Huynh T.D.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 563-1056 FROM N.A.
 RX MEDLINE=89380219; PubMed=2528543;
 RA Antonsson P., Heinegaard D., Oldberg A.;
 RT "The keratan sulfate-enriched region of bovine cartilage proteoglycan consists of a consecutively repeated hexapeptide motif.";
 RL J. Biol. Chem. 264:16170-16173(1989).
 RN [3]
 RP SEQUENCE OF 1609-2113 AND 2151-2364 FROM N.A.
 RX MEDLINE=87270630; PubMed=3111460;
 RA Oldberg A., Antonsson P., Heinegaard D.;
 RT "The partial amino acid sequence of bovine cartilage proteoglycan, deduced from a cDNA clone, contains numerous Ser-Gly sequences arranged in homologous repeats.";
 RL Biochem. J. 243:255-259(1987).
 RN [4]
 RP SEQUENCE OF 2114-2150 FROM N.A.
 RC TISSUE=Cartilage;
 RX MEDLINE=93352525; PubMed=8349621;
 RA Fuellep C., Waicz E., Vailon M., Glant T.T.;
 RT "Expression of alternatively spliced epidermal growth factor-like domains in aggregates of different species. Evidence for a novel module.";
 RL J. Biol. Chem. 268:17377-17383(1993).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=85027710; PubMed=6489519;
 RA Perin J.-P., Bonnet F., Jolles J., Jolles P.;
 RT "Sequence data concerning the protein core of the cartilage proteoglycan monomers. Characterization of a sequence allowing the synthesis of an oligonucleotide probe.";
 RL FEBS Lett. 176:37-42(1984).
 RN [6]

RP PARTIAL SEQUENCE.
 RX MEDLINE=87005253; PubMed=3530809;
 RA Perin J.-P., Bonnet F., Jolles P.;
 RT "Structural relationship between link proteins and proteoglycan monomers.";
 RL FEBS Lett. 206:73-77(1986).
 CC -!- FUNCTION: This proteoglycan is a major component of extracellular matrix of cartilaginous tissues. A major function of this protein is to resist compression in cartilage. It binds avidly to hyaluronic acid via an amino-terminal globular region. May play a regulatory role in the matrix assembly of the cartilage.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P13608-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P13608-2; Sequence=VSP_003072;
 CC -!- DOMAIN: Two globular domains, G1 and G2, comprise the amino terminus of the proteoglycan, while another globular region, G3, makes up the C-terminus. G1 contains link domains and thus consists of three disulfide-bonded loop structures designated as the A, B, B' motifs. G2 is similar to G1. The keratan sulfate (KS) and the chondroitin sulfate (CS) attachment domains lie between G2 and G3.
 CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND O-LINKED (ABOUT 40) OLIGOSACCHARIDES.
 CC -!- PTM: THE KERATAN SULFATE CONTENTS DIFFER CONSIDERABLY BETWEEN ADULT AND FETAL BOVINE PROTEOGLYCANS.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 4 link domains.
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

 DR EMBL; U76615; AAB38524.1; -;
 DR EMBL; L07053; -; NOT_ANNOTATED_CDS.
 DR PIR; A34234; A39808.
 DR PIR; T42630; T42630.
 DR HSPSP; P08709; IBF9.
 DR InterPro; IPR002353; AntifreezeZell.
 DR InterPro; IPR000152; Asx_hydroxyl_s.
 DR InterPro; IPR000742; EGF-2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR003324; SGXSG.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00059; Lectin_c; 1.
 DR Pfam; PF02339; SGXSG; 61.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 4.
 DR PRINTS; PR00356; ANTIFREEZEII.
 DR PRINTS; PR01265; LINKMODULE.
 DR ProDom; PD000918; Link; 4.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.

DR SMART; SM00445; LINK; 4.
DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
DR PROSITE; PS0041; C-TYPE LECTIN 2; 1.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS00026; EGF 3; 1.
DR PROSITE; PS01187; EGF CA; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS00290; IG-MHC; FALSE_NEG.
DR PROSITE; PS01241; LINK; 4.
KW Glycoprotein; Proteoglycan; Lectin; Signal; Sushi; EGF-like domain;
KW Calcium; Alternative splicing; Repeat; Immunoglobulin domain.
FT SIGNAL 1 16
FT CHAIN 17 2364
FT DOMAIN 25 147
FT DOMAIN 170 247
FT DOMAIN 268 349
FT DOMAIN 504 581
FT DOMAIN 602 683
FT DOMAIN 774 907
FT DOMAIN 1433 2112
FT DOMAIN 2113 2149
FT DOMAIN 2114 2364
FT DOMAIN 2161 2276
FT DOMAIN 2280 2338
FT DISULFID 51 133
FT DISULFID 175 246
FT DISULFID 199 220
FT DISULFID 273 348
FT DISULFID 297 318
FT DISULFID 509 580
FT DISULFID 533 554
FT DISULFID 607 682
FT DISULFID 631 652
FT DISULFID 2117 2128
FT DISULFID 2182 2274
FT DISULFID 2250 2266
FT DISULFID 2281 2324
FT DISULFID 2310 2337
FT CARBOHYD 126 126
FT CARBOHYD 239 239
FT CARBOHYD 333 333
FT CARBOHYD 387 387
FT CARBOHYD 611 611
FT CARBOHYD 667 667
FT VARSPLIC 2114 2150
SQ SEQUENCE 2364 AA; 246359 MW; 6FF83763420C3D4C CRC64;
Query Match 53.3%; Score 40; DB 1; Length 2364;
Best Local Similarity 53.3%; Pred. No. 3.6e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 AGLPAVGLSPGEQE 15
Db :||| |||||
971 SGLPVEGLSPGEE 985

Search completed: May 7, 2004, 12:34:34
Job time : 6.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:26:40 ; Search time 28.05 Seconds
(without alignments)
168.726 Million cell updates/sec

Title: US-09-786-214A-15

Perfect score: 75

Sequence: 1 AGLPAVVLSPGEQE 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: *
1: sp archaea: *
2: sp bacteria: *
3: sp fungi: *
4: sp human: *
5: sp invertebrate: *
6: sp mammal: *
7: sp mhc: *
8: sp organelle: *
9: sp phage: *
10: sp plant: *
11: sp rodent: *
12: sp virus: *
13: sp vertebrate: *
14: sp unclassified: *
15: sp virus: *
16: sp bacteriaph: *
17: sp archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	73.3	41	11 Q8K408	Q8K408 rattus norv
2	50	66.7	381	16 Q9RL15	Q9RL15 deinococcus
3	46	61.3	150	16 Q8PE55	Q8PE55 xanthomonas
4	46	61.3	422	16 Q9KZB0	Q9KZB0 streptomyces
5	44	58.7	335	16 Q8ZRR7	Q8ZRR7 streptomyces
6	44	58.7	383	16 Q7V8U3	Q7V8U3 prochlorococcus
7	44	58.7	399	16 Q8DKG5	Q8DKG5 synecococcus
8	43.5	58.0	614	16 Q8V280	Q8V280 ralstonia s
9	43	57.3	579	16 Q8NEJ7	Q8NEJ7 corynebacte
10	43	57.3	582	16 Q8FQM6	Q8FQM6 corynebacte
11	43	57.3	821	17 Q9HPR8	Q9HPR8 halobacteri
12	42.5	56.7	305	16 Q89FB7	Q89FB7 bradyrhizob
13	42	56.0	254	16 Q89BH1	Q89BH1 bradyrhizob
14	42	56.0	326	1 Q9UXF0	Q9UXF0 methanoblob
15	42	56.0	574	2 Q54756	Q54756 synecococc
16	42	56.0	660	9 Q7Y4W9	Q7Y4W9 bacterioph

17	42	56.0	673	13 Q7T2X8	Q7T2X8 gallus gall
18	42	56.0	1541	6 Q8HXL3	Q8HXL3 sus scrofa
19	41	54.7	143	4 Q8NAW2	Q8NAW2 homo sapien
20	41	54.7	164	16 Q99QD9	Q99QD9 caulobacter
21	41	54.7	177	17 Q9YAM1	Q9YAM1 aeropyrum p
22	41	54.7	214	16 Q9I2S0	Q9I2S0 pseudomonas
23	41	54.7	222	17 Q978T4	Q978T4 thermoplasm
24	41	54.7	239	16 Q89JT2	Q89JT2 bradyrhizob
25	41	54.7	438	17 Q8QOK1	Q8QOK1 methanosarc
26	41	54.7	448	16 Q8PML2	Q8PML2 xanthomonas
27	41	54.7	520	3 Q8N1L6	Q8N1L6 saccharomyc
28	41	54.7	520	3 Q8N1L6	Q8N1L6 homo sapien
29	41	54.7	11096	2 Q9L4W3	Q9L4W3 streptomyce
30	40	53.3	85	6 Q7786	Q7786 canis famli
31	40	53.3	130	17 Q9YFJ5	Q9YFJ5 aeropyrum p
32	40	53.3	154	17 Q8ZJX3	Q8ZJX3 pyrobaculum
33	40	53.3	242	16 Q7WCX1	Q7WCX1 bordetella
34	40	53.3	261	16 Q8NOA2	Q8NOA2 corynebacte
35	40	53.3	278	2 Q9F0Y4	Q9F0Y4 xanthomonas
36	40	53.3	298	16 Q7W5D6	Q7W5D6 bordetella
37	40	53.3	298	16 Q7V7I2	Q7V7I2 bordetella
38	40	53.3	331	16 Q7WEG7	Q7WEG7 bordetella
39	40	53.3	331	16 Q7W350	Q7W350 bordetella
40	40	53.3	331	16 Q7W032	Q7W032 bordetella
41	40	53.3	355	10 Q947A7	Q947A7 nitellopsis
42	40	53.3	357	4 Q8NEX1	Q8NEX1 homo sapien
43	40	53.3	358	4 Q9NWD0	Q9NWD0 homo sapien
44	40	53.3	358	10 Q9FUG7	Q9FUG7 fragaria an
45	40	53.3	358	12 Q91201	Q91201 havana toma

ALIGNMENTS

RESULT 1
Q8K408
ID Q8K408 PRELIMINARY; PRT; 41 AA.
AC Q8K408;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Truncated macrophage colony stimulating factor.
GN CSF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEW.tl;
RX MEDLINE=22069908; PubMed=12074592;
RA Dobbins D.E.; Sood R.; Hashimoto A.; Hansen C.T.; Wilder R.L.;
RA Remmers E.F.;
RT "Mutation of macrophage colony stimulating factor (Csfl) causes osteopetrosis in the tl rat."
RL Biochem. Biophys. Res. Commun. 294:1114-1120(2002).
DR EMBL; AF514357; AAM54137.1; -
SQ SEQUENCE 41 AA; 4178 MW; ID342C19BD18AA41 CRC64;
Query Match 73.3%; Score 55; DB 11; Length 41;
Best Local Similarity 78.6%; Pred. No. 0.1;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 GLPAVVLSPGEQE 15
Db 20 GLPAAGLSPREQE 33
RESULT 2
Q9RL15
ID Q9RL15 PRELIMINARY; PRT; 381 AA.
AC Q9RL15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Cytochrome P450.
 GN DR2473.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RL / ATCC 13939 / DSM 20539 / NCIB 9279;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Aravind K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RL radiodurans RL";
 RL Science 286:1571-1577(1999)
 CC -|- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AE002076; AAF12016.1; -.
 DR FIR; F75270; F75270.
 DR TIGR; DR2473; -.
 DR GO; GO:0004497; F:monooxygenase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME P450; 1.
 DR Heme; Monooxygenase; Oxidoreductase; Complete proteome.
 SQ SEQUENCE 381 AA; 41940 MW; F191EA69F1797B53 CRC64;

 Query Match 66.7%; Score 50; DB 16; Length 381;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 2 GLPAVVGGLSP 11
 DB 51 GLPAVVGGLSP 60

 RESULT 3
 Q8PE55 PRELIMINARY; PRT; 150 AA.
 ID Q8PE55
 AC Q8PE55;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Deoxycytidylate deaminase.
 GN XCC0126.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPPB 528;
 RX MEDLINE=4202145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
 RA Cicalalli R.M.B., Coutinho L.B., Cursino-Santos J.R., El-Dorri H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.H.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RL Nature 417:459-463(2002).
 DR EMBL; AE012107; AAM39445.1; -.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR002125; dCMP/cyt_deam.
 DR Pfam; PF00383; dCMP_cyt_deam; 1.
 KW Complete proteome.
 SQ SEQUENCE 150 AA; 16201 MW; 63D9C17D44DC9B43 CRC64;

 Query Match 61.3%; Score 46; DB 16; Length 150;
 Best Local Similarity 60.0%; Pred. No. 12;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

 QY 1 AGLPAVVGGLSPGEQ 15
 DB 104 AGIKRWVALAFGESE 118

 RESULT 4
 Q3KZB0 PRELIMINARY; PRT; 462 AA.
 ID Q3KZB0
 AC Q3KZB0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative gntR-family regulator.
 GN SC04836 OR SC5G8.04.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1302;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Saunders D.C., Harris D.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wierzbicka A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RL coelicolor A3(2)";
 RL Nature 417:141-147(2002).
 CC -|- SIMILARITY: BELONGS TO THE GNTR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.

```

DR EMBL; AL939121; CAB89055.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000524; HTH GntR.
DR InterPro; IPR000408; Reg_chir_condens.
DR Pfam; PF00392; gntR; 1.
DR PRINTS; PR00035; HTHGNTR.
DR SMART; SM00345; HTH GNTN; 1.
DR PROSITE; PS00626; RC1.2; 1.
KW DNA-binding; Transcription regulation; Complete proteome.
SQ SEQUENCE 462 AA; 48831 MW; DFCECDAGSLEI13317 CRC64;

Query Match 61.3%; Score 46; DB 16; Length 462;
Best Local Similarity 60.0%; Pred. No. 41;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLPAVVGLSPGEQ 15
DB 392 AGLHVLGLPPGTEQ 406

RESULT 5
Q82RR7 PRELIMINARY; PRT; 335 AA.
AC Q82RR7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative terpene cyclase.
GN TPC1 OR SAV76.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005021; BAC67785.1; -.
DR InterPro; IPR008949; Terpenoid_synth.
KW Complete proteome.
SQ SEQUENCE 335 AA; 36480 MW; 49B8477E2D52666F CRC64;

Query Match 58.7%; Score 44; DB 16; Length 335;
Best Local Similarity 72.7%; Pred. No. 62;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLPAVVGLSPG 12
DB 8 GLPAPAGISPG 18

RESULT 6
Q7V8U3 PRELIMINARY; PRT; 383 AA.
ID Q7V8U3
AC Q7V8U3;

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phage integrase.
GN PMT0234.
OS Prochlorococcus marinus (strain MIT 9313).
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=74547;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825698; PubMed=12917642;
RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,
RA Akgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb E.A., Zinser E.R., Chisholm S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation.";
RL Nature 424:1042-1047(2003).
DR EMBL; BX572095; CAE20409.1; -.
KW Complete proteome.
SQ SEQUENCE 383 AA; 43234 MW; CF956FE676F2E59B CRC64;

Query Match 58.7%; Score 44; DB 16; Length 383;
Best Local Similarity 75.0%; Pred. No. 71;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLPAVVGLSPGE 13
DB 193 GLPASVRLTPGE 204

RESULT 7
Q8DKG5 PRELIMINARY; PRT; 399 AA.
AC Q8DKG5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tlr0894 protein.
GN TLR0894.
OS Synecchococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimp S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005372; BAC08446.1; -.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 399 AA; 45259 MW; 61F60CE5B22A3F94 CRC64;

Query Match 58.7%; Score 44; DB 16; Length 399;
Best Local Similarity 64.3%; Pred. No. 74;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPAVVGLSPGEQ 15
DB 273 GLPFFVKGVSFYEQQ 286

RESULT 8
Q8Y280

```

```

ID Q8Y280 PRELIMINARY; PRT; 614 AA.
AC
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Probable ATP-binding transport ABC transporter protein.
GN RSC0456 OR RS0444.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choine N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502 (2002).
DR EMBL; AL646059; CAD13984.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR001623; DnaJ N.
DR Pfam; PF00005; ABC tran.1
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
KW Complete proteome.
SQ SEQUENCE 614 AA; 59240 MW; E29335B85872142 CRC64;

Query Match 58.0%; Score 43.5; DB 16; Length 614;
Best Local Similarity 45.8%; Pred. No. 1.4e+02;
Matches 11; Conservative 2; Mismatches 0; Indels 11; Gaps 1;

QY 3 LPAVVG-----LSPGQE 15
Db 501 LPALVGLDEVSNWLSRLSPGQQ 524

RESULT 9
Q8NRJ7 PRELIMINARY; PRT; 579 AA.
AC
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Sulfate permease and related transporters (MFS superfamily).
GN CGL1051.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005277; BAB98444.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008271; F:sulfate porter activity; IEA.
DR GO; GO:0008272; P:sulfate transport; IEA.

```

```

DR InterPro; IPR006162; Ppantne S.
DR InterPro; IPR002645; STAS.
DR InterPro; IPR001902; Sulph_transpt.
DR Pfam; PF01740; STAS; 1.
DR Pfam; PF00916; Sulfate_transp; 1.
DR TIGRfams; TIGR00815; sulp; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
DR PROSITE; PS00801; STAS; 1.
KW Complete proteome.
SQ SEQUENCE 579 AA; 61660 MW; B70EA8BBE0C4C100 CRC64;

Query Match 57.3%; Score 43; DB 16; Length 579;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGLPAVVGL 9
Db 48 AGLPAVVGL 56

RESULT 10
Q8FQM6 PRELIMINARY; PRT; 582 AA.
AC Q8FQM6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Putative sulfate transport protein.
GN CE1093.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005217; BAC17903.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008271; F:sulfate porter activity; IEA.
DR GO; GO:0008272; P:sulfate transport; IEA.
DR InterPro; IPR002645; STAS.
DR InterPro; IPR001902; Sulph_transpt.
DR Pfam; PF01740; STAS; 1.
DR TIGRfams; TIGR00815; sulp; 1.
DR PROSITE; PS00801; STAS; 1.
KW Complete proteome.
SQ SEQUENCE 582 AA; 62545 MW; CA67CFCAB304AE58 CRC64;

Query Match 57.3%; Score 43; DB 16; Length 582;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGLPAVVGL 9
Db 51 AGLPAVVGL 59

RESULT 11
Q9HPR8 PRELIMINARY; PRT; 821 AA.
AC Q9HPR8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE DNA helicase.
GN HEL OR VNG1501G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).

```

```

OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Melti R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Iseebarger T.A., Beck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RA "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AR005064; AAC19799.1; -.
DR PIR; C84304; C84304.
DR GO; GO:0004386; F:Helicase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001478; PDZ.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00228; PDZ; 1.
DR Helicase; Complete proteome.
KW Helicase; Complete proteome.
SQ SEQUENCE 821 AA; 89848 MW; C454C76B984A5702 CRC64;

Query Match
Best Local Similarity 57.3%; Score 43; DB 17; Length 821;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AVVGLSPGSEQ 14
|||:|||||
DB 326 AVVGLSPGSEQ 335

RESULT 12
Q89FB7 PRELIMINARY; PRT; 305 AA.
AC Q89FB7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Dihydropicolinate synthase.
GN DAPA OR BLR6784.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=USDA 110;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Ideawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RA "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AF005959; BAC52049.1; -.
DR InterPro; IPR002220; DHDP.
DR Pfam; PF00701; DHDP; 1.
DR PRINTS; PR00146; DHFICNTASE.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 305 AA; 32247 MW; EB02891D5EE5A4D8 CRC64;

Query Match
Best Local Similarity 56.7%; Score 42.5; DB 16; Length 305;
Matches 11; Conservative 2; Mismatches 2; Indels 9; Gaps 1;

QY 1 AGUPAVV-----GLSPGSEQ 15
|||:|||||
DB 326 AVVGLSPGSEQ 335

Query Match
Best Local Similarity 56.0%; Score 42; DB 16; Length 254;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPAAVGLSPGSEQ 13
|||:|||||
DB 131 GLPAAVGLSPGSEQ 142

RESULT 14
Q9UXP0 PRELIMINARY; PRT; 326 AA.
AC Q9UXP0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F420-dependent N5,N10-methylene-tetrahydromethanopterin reductase,
DE putative.
GN PFDA.
OS Methanobolus tindarius.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanobolus.
OX NCBI_TaxID=2221;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DSM 2278;
RA Westerberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA Gottschalk G., Blaut M.;
RA "The F420H2-dehydrogenase from Methanobolus tindarius: Cloning of the
f420 operon and expression of the genes in Escherichia coli.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011519; CAB56639.1; -.
DR PIR; T45226; T45226.
DR InterPro; IPR002103; Bac_luciferase.
DR Pfam; PF00236; bac_luciferase; 1.
SQ SEQUENCE 326 AA; 34043 MW; 16F3AB9733A45D82 CRC64;

Query Match
Best Local Similarity 56.0%; Score 42; DB 1; Length 326;
Matches 11; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 AGUPAVV-----GLSPGSEQ 15
|||:|||||
DB 41 AGUPAVV-----GLSPGSEQ 154

```

RESULT 13

```

Q89BH1 PRELIMINARY; PRT; 254 AA.
AC Q89BH1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BL18182 protein.
GN BL18182.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=USDA 110;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Ideawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RA "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AF005964; BAC53447.1; -.
DR InterPro; IPR002559; Transposase_11.
DR Pfam; PF01609; Transposase_11; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 254 AA; 29160 MW; 76C1ED8F28D5C758 CRC64;

Query Match
Best Local Similarity 66.7%; Score 42; DB 16; Length 254;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPAAVGLSPGSEQ 13
|||:|||||
DB 131 GLPAAVGLSPGSEQ 142

```

RESULT 14

```

Q9UXP0 PRELIMINARY; PRT; 326 AA.
AC Q9UXP0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F420-dependent N5,N10-methylene-tetrahydromethanopterin reductase,
DE putative.
GN PFDA.
OS Methanobolus tindarius.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanobolus.
OX NCBI_TaxID=2221;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DSM 2278;
RA Westerberg D.J., Braune A., Ruppert C., Mueller V., Herzberg C.,
RA Gottschalk G., Blaut M.;
RA "The F420H2-dehydrogenase from Methanobolus tindarius: Cloning of the
f420 operon and expression of the genes in Escherichia coli.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011519; CAB56639.1; -.
DR PIR; T45226; T45226.
DR InterPro; IPR002103; Bac_luciferase.
DR Pfam; PF00236; bac_luciferase; 1.
SQ SEQUENCE 326 AA; 34043 MW; 16F3AB9733A45D82 CRC64;

Query Match
Best Local Similarity 56.0%; Score 42; DB 1; Length 326;
Matches 11; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 AGUPAVV-----GLSPGSEQ 15
|||:|||||
DB 41 AGUPAVV-----GLSPGSEQ 154

```


Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLPAVVGGLSPGEG 14
:|:|:|:|:|:|
Db 84 SGGRAILGLGPGEG 97

RESULT 15

Q54756 PRELIMINARY; PRT; 574 AA.
AC Q54756;
DT 01-NOV-1996 (TREMELREL. 01, Created)
DT 01-JUN-1998 (TREMELREL. 06, Last sequence update)
DT 01-OCT-2003 (TREMELREL. 25, Last annotation update)
DE Similar to plant sulfate transporter.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7942;
RA Phung L.T., Haselkorn R.;
RT "Gene encoding biotin carboxylase subunit of acetyl-CoA carboxylase
from cyanobacterium Synechococcus sp. PCC 7942.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59234; AAB88215.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0008271; F:sulfate porter activity; IEA.
DR GO; GO:0008272; F:sulfate transport; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR001902; Sulph_transpt.
DR Pfam; PF01740; STAS; 1.
DR Pfam; PF00916; Sulfate_transp; 1.
DR TIGRFAMs; TIGR00815; sulp; 1.
DR PROSITE; PS0801; STAS; 1.
SQ SEQUENCE 574 AA; 62328 MW; 7BDECD65C67BBB00 CRC64;

Query Match 56.0%; Score 42; DB 2; Length 574;
Best Local Similarity 88.9%; Pred. No. 2.3e+02;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGLPAVVGGL 9
:|:|:|:|:|:|
Db 50 AGLPAVVGGL 58

Search completed: May 7, 2004, 12:37:57
Job time : 29.2167 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 12:19:40 ; Search time 41.85 Seconds
(without alignments)
101.272 Million cell updates/sec

Title: US-09-786-214A-15
Perfect score: 75
Sequence: 1 AGLPAWGLSPGEQE 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	75	100.0	15	3	AAy84269	Peptide d
2	75	100.0	20	3	AAy84265	Truncated
3	75	100.0	25	3	AAy84264	Peptide o
4	65	86.7	13	3	AAy84268	Peptide d
5	65	86.7	14	3	AAy84266	Peptide d
6	61	81.3	13	3	AAy84267	Peptide d
7	45	60.0	234	4	AAy84268	Human imm
8	44	58.7	116	4	ABG24038	Novel hum
9	43	57.3	473	4	AAU40252	Propionib
10	43	57.3	473	6	ABM36771	Propionib
11	43	57.3	565	4	ABY76817	Propionib
12	43	57.3	579	4	AAy84267	Coryneb
13	42	56.0	470	6	AAy84268	C glutami
14	42	56.0	475	6	AAE34724	Streptomy
15	42	56.0	475	6	AAE34732	Streptomy
16	41	54.7	106	5	ABJ10397	Mutant an
17	41	54.7	106	5	ABJ10395	Mutant an
18	41	54.7	143	5	ABD21677	Human sec
19	41	54.7	143	7	ADB64633	Human pro
20	41	54.7	154	4	AAm25921	Human pro
21	41	54.7	241	5	AAm48925	scfv anti
22	41	54.7	294	1	AAp82484	Tropoelas
23	41	54.7	370	2	AAy14924	Amino aci
24	41	54.7	370	6	ABp70895	Mycobacte
25	41	54.7	372	5	ABB73530	M vaccae

26	41	54.7	396	3	AAy84269	Gene 15 h
27	41	54.7	530	7	ADD49105	Human NOV
28	41	54.7	549	7	ADD49091	Human NOV
29	41	54.7	579	5	ABG70277	Human Gly
30	41	54.7	579	5	ABG97356	Human GPC
31	41	54.7	579	6	ABR39111	Human GPC
32	41	54.7	579	7	ADD49087	Human NOV
33	41	54.7	579	7	ADD49107	Human NOV
34	41	54.7	579	7	ADD49089	Human NOV
35	41	54.7	592	7	ADD49099	Human NOV
36	41	54.7	774	3	AAy84269	Human Lys
37	41	54.7	774	5	ABM00077	Human Lys
38	41	54.7	774	6	ABM07649	Human Lys
39	41	54.7	774	6	ABU03509	Angiogene
40	41	54.7	774	6	ABU57639	Different
41	41	54.7	774	6	ABR41070	Human MAP
42	40	53.3	1196	4	AAE10129	Streptomy
43	40	53.3	119	4	AAU51147	Propionib
44	40	53.3	132	3	ABM47666	Propionib
45	40	53.3	215	5	AAy84269	Arabidops
					ABp70895	Human sec

ALIGNMENTS

RESULT 1
AAy84269
ID AAy84269 standard; peptide; 15 AA.
XX
AC AAy84269;
XX
DT 12-JUL-2000 (first entry)
XX
DE Peptide derived from macrophage colony stimulating gene alternative ORF.
XX
KW tumour rejection antigen; macrophage colony stimulating gene;
KW macrophage-colony stimulating factor; antigen presenting cell;
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200013699-A1.
XX
PD 16-MAR-2000.
XX
PF 03-SEP-1999; 99WO-US020344.
XX
PR 04-SEP-1998; 98US-0099077P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;
XX
DR WPI; 2000-256859/22.
XX
PT Isolated polypeptide used to treat subjects having a disorder
PT characterized by expression of alternative open reading frame macrophage-
PT colony stimulating factor comprises 25 amino acid residue sequence.
XX
XX Example 2; Page 40; 74pp; English.
XX
CC The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying

KW macrophage-colony stimulating factor; antigen presenting cell;
 XX human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

OS Synthetic.
 OS Homo sapiens.

FN WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 13 AA;

Query Match 86.7%; Score 65; DB 3; Length 13;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LPAVGLSPGEQE 15
 Db 1 LPAVGLSPGEQE 13

RESULT 5
 AAY84266
 ID AAY84266 standard; peptide; 14 AA.

AC AAY84266;

XX 12-JUL-2000 (first entry)

DE Peptide derived from macrophage colony stimulating gene alternative ORF.

XX tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Synthetic.
 OS Homo sapiens.

XX WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX

PR 04-SEP-1998; 98US-0099077P.
 XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

XX Claim 2; Page 39; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
 CC rejection antigen precursor encoded by an alternative open reading frame
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
 CC the alternative ORF of macrophage-colony stimulating factor, when
 CC presented by an antigen presenting cell having a human leukocyte antigen
 CC (HLA) class I molecule, effectively induce the activation and
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
 CC acids derived from the alternate ORF of macrophage-colony stimulating
 CC factor are useful for enriching selectively a population of T lymphocytes
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
 CC characterized by expression of the polypeptide, and for identifying
 CC functional variants and mimetics

XX Sequence 14 AA;

Query Match 86.7%; Score 65; DB 3; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.0017;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LPAVGLSPGEQE 15
 Db 1 LPAVGLSPGEQE 13

RESULT 6
 AAY84267
 ID AAY84267 standard; peptide; 13 AA.

AC AAY84267;

XX 12-JUL-2000 (first entry)

DE Peptide derived from macrophage colony stimulating gene alternative ORF.

XX tumour rejection antigen; macrophage colony stimulating gene;
 KW macrophage-colony stimulating factor; antigen presenting cell;
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.

XX Synthetic.
 OS Homo sapiens.

XX WO200013699-A1.

XX 16-MAR-2000.

XX 03-SEP-1999; 99WO-US020344.

XX 04-SEP-1998; 98US-0099077P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Probst-Kepper M, Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-256859/22.

XX Isolated polypeptide used to treat subjects having a disorder
 PT characterized by expression of alternative open reading frame macrophage-
 PT colony stimulating factor comprises 25 amino acid residue sequence.

PS Example 2; Page 40; 74pp; English.

XX The present sequence represents a peptide which is derived from a tumour
CC rejection antigen precursor encoded by an alternative open reading frame
CC (ORF) of human macrophage colony stimulating gene. Peptides derived from
CC the alternative ORF of macrophage-colony stimulating factor, when
CC presented by an antigen presenting cell having a human leukocyte antigen
CC (HLA) class I molecule, effectively induce the activation and
CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic
CC acids derived from the alternate ORF of macrophage-colony stimulating
CC factor are useful for enriching selectively a population of T lymphocytes
CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder
CC characterized by expression of the polypeptide, and for identifying
CC functional variants and mimetics

XX Sequence 13 AA;

Query Match 81.3%; Score 61; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 PAVVGLSPGEQ 15
Db 1 PAVVGLSPGEQ 12

RESULT 7

AAB36208
ID AAB36208 standard; protein; 234 AA.

AC AAB36208;

DT 15-FEB-2001 (first entry)

XX Human immune system associated protein HISAP-6.

XX Human; immune system associated protein; HISAP-6; immune disorder;
KW infection; autoimmune disease; cancer.

XX Homo sapiens.

PN US6135941-A.

XX 24-OCT-2000.

XX 27-MAR-1998; 98US-00049672.

XX 27-MAR-1998; 98US-00049672.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;

XX Hillman JL, Au-Young J;

XX WPI: 2001-030926/04.

XX N-PSDB; AAC66524.

XX New human immune system associated proteins (HISAP) and polynucleotides
PT encoding the HISAP, useful for diagnosing, treating or preventing immune
PT or cell proliferative disorders or infections.

XX Claim 1; Col 59-60; 54pp; English.

XX The present invention provides the coding and protein sequences for a
CC number of human immune system associated proteins (HISAPs). These can be
CC used in the diagnosis and treatment of various autoimmune disorders,
CC infections and cell proliferation diseases. The diseases include AIDS,
CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,
CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia
CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus
CC erythematosus, arteriosclerosis, cirrhosis and cancer

XX Sequence 234 AA;

Query Match

Best Local Similarity 60.0%; Score 45; DB 4; Length 234;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PAVVGLSPGEQ 14

Db 28 PAVVGLSPGER 38

RESULT 8

ABG24038

ID ABG24038 standard; protein; 116 AA.

XX ABG24038;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #24029.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB; AAS88225.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 54397; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG03077 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 116 AA;

Query Match

58.7%; Score 44; DB 4; Length 116;

Best Local Similarity 64.3%; Pred. No. 35;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPAVVGLSPGQR 15
|||:|||||:
Db 90 GLPXRLGSLFVERE 103

RESULT 9
AAU40252
ID AAU40252 standard; protein; 473 AA.
XX
AC AAU40252;
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #1148.
XX
KW SAPHO syndrome, synovitis; acne; pustulosis; hypertyosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
FN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59511.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Example 1; SEQ ID NO 1447; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertyosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 473 AA;

Best Local Similarity 63.6%; Pred. No. 2.3e+02;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLPAVVGLSPG 12
|||:|||||:
Db 446 GLPAIIGLAAG 456

RESULT 10
ABM36771
ID ABM36771 standard; protein; 473 AA.
XX
AC ABM36771;
XX
DT 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes predicted ORF-encoded polypeptide #1447.
XX
KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
XX immunostimulant; immune response; vaccine.
XX
OS Propionibacterium acnes.
XX
FN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978925.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallie-Douglass J;
XX
DR WPI; 2003-381789/36.
DR N-PSDB; ACF64440.
XX
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Example 1; SEQ ID NO 1447; 1481pp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 473 AA;
 Query Match 57.3%; Score 43; DB 6; Length 473;
 Best Local Similarity 63.6%; Pred. No. 2.3e+02;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 GLPAVGLSPG 12
 DB 446 GLPAIIGLAAG 456
 RESULT 11
 AAB76817
 ID AAB76817 standard; protein; 565 AA.
 AC AAB76817;
 DT 11-APR-2001 (first entry)
 DE Corynebacterium glutamicum MCT protein SEQ ID NO:616.
 XX
 KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
 KW membrane construction and membrane transport protein; petroleum spill;
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
 KW identification; microorganism; fine chemical production; transformation;
 KW genome mapping; genetic engineering.
 XX OS
 XX Corynebacterium glutamicum.
 PN WO200100805-A2.
 XX PD 04-JAN-2001.
 XX PF 23-JUN-2000; 2000WO-IB000926.
 XX PR 25-JUN-1999; 93US-0141031P.
 PR 08-JUL-1999; 93DE-0103145A.
 PR 08-JUL-1999; 93DE-0103147B.
 PR 08-JUL-1999; 93DE-01031563.
 PR 09-JUL-1999; 93DE-01032122.
 PR 09-JUL-1999; 93DE-01032124.
 PR 09-JUL-1999; 93DE-01032125.
 PR 09-JUL-1999; 93DE-01032128.
 PR 09-JUL-1999; 93DE-01032180.
 PR 09-JUL-1999; 93DE-01032182.
 PR 09-JUL-1999; 93DE-01032190.
 PR 09-JUL-1999; 93DE-01032191.
 PR 09-JUL-1999; 93DE-01032205.
 PR 09-JUL-1999; 93DE-01032212.
 PR 09-JUL-1999; 93DE-01032227.
 PR 09-JUL-1999; 93DE-01032228.
 PR 09-JUL-1999; 93DE-01032229.
 PR 09-JUL-1999; 93DE-01032230.
 PR 14-JUL-1999; 93DE-01032927.
 PR 14-JUL-1999; 93DE-01033005.
 PR 14-JUL-1999; 93DE-01033006.
 PR 27-AUG-1999; 93DE-01040764.
 PR 27-AUG-1999; 93DE-01040765.
 PR 27-AUG-1999; 93DE-01040766.
 PR 27-AUG-1999; 93DE-01040830.
 PR 27-AUG-1999; 93DE-01040831.
 PR 27-AUG-1999; 93DE-01040832.
 PR 27-AUG-1999; 93DE-01040833.
 PR 31-AUG-1999; 93DE-01041378.
 PR 31-AUG-1999; 93DE-01041379.
 PR 31-AUG-1999; 93DE-01041395.
 PR 03-SEP-1999; 93DE-01042077.
 PR 03-SEP-1999; 93DE-01042078.
 PR 03-SEP-1999; 93DE-01042079.
 PR 03-SEP-1999; 93DE-01042088.
 XX (BADI) BASF AG.

XX PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 XX DR WPI; 2001-071486/08.
 XX N-PSDB; AAF68050.
 XX PT Corynebacterium glutamicum nucleic acids encoding membrane construction
 PT and membrane transport proteins or their portions, useful for typing or
 PT identifying C. glutamicum or related bacteria, and as markers for
 PT transformation.
 XX PS Claim 20; Page 1030-1031; 1119pp; English.
 XX CC AAB767743 to AAF68080 encode the Corynebacterium glutamicum membrane
 CC construction and membrane transport (MCT) proteins given in AAB76510 to
 CC AAB76847. The MCT nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 CC identification of C. glutamicum or related bacteria, as reference points
 CC for mapping C. glutamicum genome, and as markers for transformation.
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an
 CC example from the present invention
 XX SQ Sequence 565 AA;
 Query Match 57.3%; Score 43; DB 4; Length 565;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGLPAVGL 9
 DB 34 AGLPAVGL 42
 RESULT 12
 AAG90917
 ID AAG90917 standard; protein; 579 AA.
 AC AAG90917;
 DT 26-SEP-2001 (first entry)
 DE C glutamicum protein fragment SEQ ID NO: 4671.
 XX KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX OS Corynebacterium glutamicum.
 XX PN EP1108790-A2.
 XX PD 20-JUN-2001.
 XX PF 18-DEC-2000; 2000EP-00127688.
 PR 16-DEC-1999; 93JP-00377484.
 PR 07-APR-2000; 2000JP-00159162.
 PR 03-AUG-2000; 2000JP-00280988.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX WPI; 2001-376931/40.
 XX N-PSDB; AAF66136.
 XX PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analyzing
 PT expression profile or pattern of a gene and identifying homologous gene.
 XX PS Claim 17; SEQ ID NO 4671; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and analysing
 CC the expression profile or expression pattern of a gene derived from
 CC coryneform bacterium, and identifying a homologue of a gene derived from
 CC coryneform bacterium. Coryneform bacteria are useful for producing amino
 CC acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described in the
 CC exemplification of the invention. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the European Patent Office
 XX
 XX Sequence 579 AA;

Query Match 57.3%; Score 43; DB 4; Length 579;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGLPAVVGL 9
 Db 48 AGLPAVVGL 56
 ||| ||| |||

RESULT 13
 AAE34724
 ID AAE34724 standard; protein; 470 AA.
 AC AAE34724;
 XX
 XX 14-MAY-2003 (first entry)
 XX Streptomyces rimosus ema3 protein.
 DE
 XX P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
 KW emamectin; insecticide; ema3 protein.
 XX
 OS Streptomyces rimosus.
 XX
 PN WO200292801-A2.
 XX
 XX 21-NOV-2002.
 XX
 PF 15-MAY-2002; 2002WO-EP005363.
 XX
 PR 16-MAY-2001; 2001US-0291149P.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
 PI Buckel TG;
 XX
 DR WPI; 2003-140280/13.
 DR N-PSDB; AAD53019.
 XX
 PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
 PT and capable of regioselectively oxidizing alcohol of avermectin useful
 PT for making emamectin from avermectin.
 XX
 PS Claim 17; Page 107-108; 157pp; English.

XX The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidising
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferrodoxins and ferrodoxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces rimosus ema3 protein
 XX
 XX Sequence 470 AA;

Query Match 56.0%; Score 42; DB 6; Length 470;
 Best Local Similarity 72.7%; Pred. No. 3.3e+02;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 LPAVVGLSPGE 13
 Db 44 LPSVGLHPGE 54
 ||| ||| |||

RESULT 14
 AAE34732
 ID AAE34732 standard; protein; 475 AA.
 AC AAE34732;
 XX
 XX 14-MAY-2003 (first entry)
 XX Streptomyces rimosus ema1 protein.
 DE
 XX P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
 KW emamectin; insecticide; ema1 protein.
 XX
 OS Streptomyces rimosus.
 XX
 PN WO200292801-A2.
 XX
 XX 21-NOV-2002.
 XX
 PF 15-MAY-2002; 2002WO-EP005363.
 XX
 PR 16-MAY-2001; 2001US-0291149P.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
 PI Buckel TG;
 XX
 DR WPI; 2003-140280/13.
 DR N-PSDB; AAD53027.
 XX
 PT Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
 PT and capable of regioselectively oxidizing alcohol of avermectin useful
 PT for making emamectin from avermectin.
 XX
 PS Claim 17; Page 121-122; 157pp; English.

XX The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidising
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferrodoxins and ferrodoxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces rimosus ema1 protein
 XX
 XX Sequence 475 AA;

Query Match 56.0%; Score 42; DB 6; Length 475;
 Best Local Similarity 72.7%; Pred. No. 3.3e+02;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 LPAVVGLSPGE 13
 Db 49 LPSVGLHPGE 59
 ||| ||| |||

RESULT 15
 AAE34729
 ID AAE34729 standard; protein; 475 AA.
 XX
 AC AAE34729;
 XX
 XX 14-MAY-2003 (first entry)

XX Streptomyces albofaciens ema8 protein.
DE
XX P450 monooxygenase; avermectin; ferrodexin; ferrodexin reductase; enzyme;
KW emamectin; insecticide; ema8 protein.
KW
XX Streptomyces albofaciens.
OS
XX WO200292801-A2.
PN
XX
XX 21-NOV-2002.
PD
XX 15-MAY-2002; 2002WO-BE005363.
PF
XX 16-MAY-2001; 2001US-0291149P.
PR
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
PA
XX Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI Buckel TG;
PI
XX WPI; 2003-140280/13.
DR
XX N-PSDB; AAD53024.
DR
XX Novel polypeptide that exhibits enzymatic activity of P450 monooxygenase
PT and capable of regioselectively oxidizing alcohol of avermectin useful
PT for making emamectin from avermectin.
PT
XX Claim 17; Page 116-117; 157pp; English.
PS
XX The present invention relates to novel proteins that exhibit an enzymatic
CC activity of P450 monooxygenase and capable of regioselectively oxidising
CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
CC The invention also relates to ferrodexins and ferrodexin reductases that
CC are active with the P450 monooxygenases. Sequences of the invention are
CC useful for producing 4-keto-avermectin from avermectin, which is useful
CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is Streptomyces albofaciens ema8 protein
XX
SQ Sequence 475 AA;

Query Match 56.0%; Score 42; DB 6; Length 475;
Best Local Similarity 72.7%; Pred. No. 3.3e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 LPAVVGLSPGE 13
Db 49 LPSYVGLHPGE 59
||: ||| |||
||: ||| |||

Search completed: May 7, 2004, 12:33:48
Job time : 43.85 secs